



ZACHODNIOPOMORSKI UNIWERSYTET TECHNOLOGICZNY W SZCZECINIE

WYDZIAŁ INFORMATYKI

mgr inż. Łukasz Kupracz

**Metody predykcji w szeregach czasowych oparte na  
sztucznej inteligencji w wybranych procesach złożonych**

*Rozprawa doktorska*

Promotor: prof. dr hab. inż. Antoni Wiliński

Szczecin 2024



*Szczególne podziękowania dla promotora niniejszej rozprawy  
prof. dr hab. inż. Antoniego Wilińskiego  
za pokazanie drogi do rozwoju naukowego, współpracę i cierpliwość.  
Dziękuję mojej żonie i dzieciom za wsparcie w trakcie prowadzenia badań naukowych oraz  
tworzenia niniejszej rozprawy.*

## Spis treści

<b>1</b>	<b>WPROWADZENIE</b> .....	<b>9</b>
1.1	CEL PRACY.....	11
1.1.1	<i>Motywacja</i> .....	11
1.2	PANDEMIA I INFLACJA A PROCESY ZŁOŻONE.....	13
1.3	SPOŁECZNE ZNACZENIE INFLACJI I JEJ PROGNOZ.....	18
1.4	WPLYW WYBRANYCH PANDEMII NA SPOŁECZEŃSTWO.....	20
1.4.1	<i>Dżuma Justyniana</i> .....	21
1.4.2	<i>Czarna śmierć</i> .....	21
1.4.3	<i>Cholera – 7 pandemii</i> .....	22
1.4.4	<i>Grypa „hiszpanka”</i> .....	23
1.4.5	<i>HIV i AIDS</i> .....	23
1.5	WPLYW PANDEMII COVID-19 NA ŚRODOWISKO NATURALNE.....	23
1.5.1	<i>Środowisko naturalne</i> .....	24
1.5.2	<i>Rozwój technologiczny</i> .....	27
1.5.3	<i>Stosowanie obostrzeń</i> .....	28
1.5.4	<i>Kwarantanna</i> .....	29
1.5.5	<i>Izolacja</i> .....	30
1.5.6	<i>Ograniczenie mobilności</i> .....	30
1.5.7	<i>Komunikacja społeczna i jakość danych w dobie kryzysu</i> .....	33
1.5.8	<i>Spójność danych statystycznych na przykładzie Polski</i> .....	34
1.6	KRAJOWE ZESPOŁY EKSPERCKIE.....	34
1.6.1	<i>Zespół modelu epidemiologicznego Interdyscyplinarnego Centrum Modelowania Uniwersytetu Warszawskiego</i> .....	35
1.6.2	<i>Zespół ds. monitorowania i prognozowania epidemii COVID-19 powołany przez Ministra Zdrowia</i> .....	36
1.6.3	<i>Departament Analiz i Strategii Ministerstwa Zdrowia</i> .....	36
1.6.4	<i>Grupa MOCOS z Politechniki Wrocławskiej</i> .....	37
1.7	POWSZECHNIE STOSOWANE METODY PREDYKCJI ROZPRZESTRZENIANIA SIĘ WIRUSÓW.....	37
1.7.1	<i>SIR i metody pochodne</i> .....	39
1.7.2	<i>Model matematyczny oparty na agentach</i> .....	46
1.7.3	<i>Model Bhatnagar’a</i> .....	48

1.7.4	<i>Model ARIMA</i> .....	51
1.7.5	<i>Deep learning</i> .....	56
1.8	INFLACJA JAKO DRUGI SPOŁECZNIE ISTOTNY CEL MODELOWANIA .....	58
1.8.1	<i>Modele inflacji i metody jej obliczania</i> .....	59
1.8.2	<i>Inflacja w państwach OECD</i> .....	63
1.8.3	<i>Inflacja w Polsce</i> .....	68
1.9	SFORMUŁOWANIE PROBLEMU BADAWCZEGO .....	70
1.10	HIPOTEZA BADAWCZA.....	71
<b>2</b>	<b>METODY PREDYKCJI ISTOTNYCH SPOŁECZNIE SZEREGÓW CZASOWYCH. ....</b>	<b>71</b>
2.1	METODA MULTIREGRESJI Z WYKORZYSTANIEM PSEUDOINWERSJI MOORE'A-PENROSE'A.....	73
2.1.1	<i>Przygotowanie i preprocesowanie danych</i> .....	75
2.1.2	<i>Zaprojektowanie modelu</i> .....	78
2.1.3	<i>Odczyt uzyskanych wyników</i> .....	85
2.1.4	<i>Pomiar efektywności</i> .....	86
2.2	METODA OPARTA NA ŁAŃCUCHACH MARKOWA.....	86
2.2.1	<i>Wprowadzenie</i> .....	86
2.2.2	<i>Predykcja liczby zakażeń</i> .....	94
2.2.3	<i>Określenie błędu APE i MAPE</i> .....	94
2.3	METODA OPARTA NA PODOBIENSTWIE DYNAMIKI ZAKAŻEŃ W N-NAJBLIŻSZYCH SĄSIADACH Z WYKORZYSTANIEM ŚREDNIEJ KROCZĄCEJ. ....	95
2.3.1	<i>Przygotowanie danych liczbowych</i> .....	95
2.3.2	<i>Analiza zbioru danych wejściowych</i> .....	100
2.3.3	<i>Walidacja i weryfikacja wyniku</i> .....	104
2.3.4	<i>Model oparty na średniej kroczącej</i> .....	108
2.3.5	<i>Predykcja dobowych skumulowanych zakażeń</i> .....	113
2.3.6	<i>Ewaluacja wyników</i> .....	114
2.4	PROGNOZA INFLACJI NA POZIOMIE KRAJU.....	116
2.4.1	<i>Metoda regresji wielomianem z optymalizacją parametrów</i> .....	116
2.4.2	<i>Metoda korekcji błędów predykcji za pomocą nieliniowej regresji</i> .....	120
2.4.3	<i>Metoda regresji wygładzania splajnem z optymalizacją parametrów</i> .....	122
<b>3</b>	<b>BADANIE DOKŁADNOŚCI ZAPREZENTOWANYCH METOD PREDYKCJI ROZPRZESTRZENIANIA SIĘ WIRUSA. ....</b>	<b>127</b>
3.1	PLAN PRAC BADAWCZYCH.....	128
3.2	BADANIE WPŁYWU GĘSTOŚCI SZCZEPIEŃ NA DOKŁADNOŚĆ PREDYKCJI.....	133
3.3	BADANIE WPŁYWU DŁUGOŚCI HORYZONTU PREDYKCJI I LICZBY SĄSIADÓW DLA OKREŚLONEGO KRAJU.....	147

3.4	BADANIE DOKŁADNOŚCI METOD PREDYKCJI W TOŻSAMYM INTERWALE CZASOWYM W ZALEŻNOŚCI OD DŁUGOŚCI HORYZONTU PREDYKCJI .....	161
3.4.1	<i>Metoda multiregresji z pseudoinwersją Moore'a-Penrose'a</i> .....	163
3.4.2	<i>Metoda oparta na n-najbliższych sąsiadach i średniej kroczącej</i> .....	168
3.4.3	<i>Metoda oparta na łańcuchach Markowa</i> .....	173
3.4.4	<i>Zbiorcze porównanie metod predykcji w zależności od horyzontu redykcji pod względem MAPE.</i>	178
3.5	WPLYW POŁOŻENIA GEOGRAFICZNEGO NA DOKŁADNOŚĆ PREDYKCJI. ....	180
3.6	BADANIE DOKŁADNOŚCI PREDYKCJI W ZALEŻNOŚCI OD FAZY PANDEMII DLA WYBRANEGO KRAJU. ....	190
<b>4</b>	<b>PROGNOZOWANIE INFLACJI - METODA OPARTA NA SKORYGOWANYM MODELU REGRESJI</b>	
	<b>200</b>	
4.1	PRZYGOTOWANIE DANYCH .....	200
4.2	ODCZYT UZYSKANYCH WYNIKÓW .....	202
4.3	POMIAR DOKŁADNOŚCI METODY .....	204
<b>5</b>	<b>KONKLUZJE</b> .....	<b>213</b>
5.1	DYSKUSJA WYNIKÓW BADAŃ .....	213
5.2	DALSZE KIERUNKI BADAŃ .....	223

*Słownik pojęć i synonimów użytych w pracy*

Pojęcie/synonim	Wyjaśnienie
ABM	Agent-based model
ACM	ACM Computing Classification System
AIC	Akaike's Information Criterion
ANN	(ang.: Artificial Neural Network), Sztuczna Sieć Neuronowa
AR	Autoregresssion
ARDL	(ang.: Autoregressive Distributed Lag), Model Autoregresji z Rozłożonymi Opóźnieniami
ARIMA	Auto-Regresive Integrated Moving Average
ARMA	Auto-Regresive Moving Average
ASF	Afrykański pomór świń
BEL	Belgia
BIC	Bayesian Information Criterion
BMA	(ang.: Bayesian Model Averaging), Bayesowskie Uśrednianie Modeli
BRA	Brazylia
CCPI	(ang.: Colombo Consumer Price Index), Indeks Cen Konsumpcyjnych w Kolombo
CHN	Chiny
COVID-19	ostra choroba zakaźna układu oddechowego wywołana zakażeniem wirusem SARS-CoV-2.
CP-ABM	Change point – Agent-based modelling
CPI	(ang.: Consumer Price Index), Indeks Cen Konsumpcyjnych
DEU	Niemcy
DFM	(ang.: Dynamic Factor Model), Dynamiczny Model Czynnikiowy
DNK	Dania
ECM	(ang.: Error Correction Model), model korekcji błędu
ECMWF	The European Centre for Medium-Range Weather Forecasts
EST	Estonia
Euro WHO	Europejskie Biuro Regionalne WHO
EV	Explained variance
EVDHM	Eigenvalue decomposition of Hankel matrix
FTSA	Functional Time Series Analysis
GCV	(ang.: Generalized Cross-Validation) Uogólniona Walidacja Krzyżowa
GDP	(ang.: Gross Domestic Product), Produkt Krajowy Brutto
GQM	Goal Question Metric
GRU	Gated recurrent units
HIV	(z ang. human immunodeficiency virus), ludzki wirus niedoboru odporności, ludzki wirus upośledzenia odporności

HMM	Hidden Markow Model
HRNN	(ang.: Hierarchical Recurrent Neural Network), Hierarchiczna Rekurencyjna Sieć Neuronowa
HUN	Węgry
ICM	Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego
IHR	International Health Regulations
IND	Indie
ISO 3166-1 alpha-3	
KNN	(ang.: K-Nearest Neighbors), K-Najbliższych Sąsiadów
LR	(ang.: Linear Regression), Regresja Liniowa
LSTM	Long Short Term Memory
LTU	Litwa
MA	Moving average
MAE	Mean Absolute Error
MAE	Mean Average Error
MAPE	Mean Absolute Percentage Error
Max Z	maksymalnym wynik testu Z (testy stosowane do identyfikacji ekstremalnych obserwacji – outlierów)
MERS	(z ang. <i>Middle East respiratory syndrome</i> ), bliskowschodni zespół oddechowy
MEX	Meksyk
MIDAS	Mixed Data Sampling
MLR	(ang.: Multiple Linear Regression), Wielokrotna Regresja Liniowa
MOCOS	MOdelling COronavirus Spread
MSE	Mean squared error
MSE	Mean Squared Error
MSSA	(ang.: Multivariate Singular Spectrum Analysis), Wielowymiarowa Analiza Jednostkowego Spektrum
MŚP	Małe i Średnie Przedsiębiorstwa
NLD	Holandia
NLP	Natural Language Processing
OECD	(ang. Organisation for Economic Co-operation and Development), Organizacja Współpracy Gospodarczej i Rozwoju
PCE	(ang.: Personal Consumption Expenditures), Wydatki na Konsumpcję Indywidualną
PCR	(z ang. Polymerase Chain Reaction) reakcja łańcuchowa polimerazy
PKB	Produkt Krajowy Brutto
POL	Polska

POZ	Podstawowa Opieka Zdrowotna
PPT	Phillips–Perron test
R <sup>2</sup>	R-squared
RF	(ang.: Random Forest), Las Losowy
RMSE	Root mean square error
RMSE	Root Mean Squared Error
RMSLE	Root Mean Squared Log Error
RUS	Rosja
RW	(ang.: Random Walk), Błądzenie Losowe
SARS	wirusowa choroba oddechowa wywoływana przez koronawirusa SARS-CoV
SARS-CoV-2	Drugi koronawirus ciężkiego ostrego zespołu oddechowego
SAU	Arabia Saudyjska
SEIAQRDT	Susceptible, Exposed, Infected, Asymptomatic, Quarantined, Recovered, Dead, Insusceptible
SEIAR	Susceptible, Exposed, Infected, Asypmtomatic, Recovered
SEIARD	Susceptible, Exposed, Infected, Asypmtomatic, Recovered, Deceased
SEIJR	Susceptible, Exposed, Infected, Isolated, Recovered
SEIR	susceptible-exposed-infectious-recovered
SIDARTHE	Susceptible, Infected, Diagnosed, Ailing, Recognized, Threatened, Healed, Extinct
SIR	susceptible-infectious-recovered
SIRD	Susceptible, Infected, Recovered, Dead
SLS	Smoothed Lag Selection
SSA	(ang.: Singular Spectrum Analysis), Analiza Jednostkowego Spektrum
SVR	Support Vector Regression
SVR	(ang.: Support Vector Regression), Regresja Wektorów Nośnych
TUR	Turcja
USA	Stany Zjednoczone
UV	Ultra violet
VAE	Variational AutoEncoder
WHO	Światowa Organizacja Zdrowia
XGBoost	(ang.: eXtreme Gradient Boosting), Ekstremalne Wzmocnienie Gradientowe
ZAF	RPA

## Streszczenie

Rozważania w rozprawie doktorskiej dotyczą dokładności metod i modeli statystycznych wykorzystywanych do prognozowania w szeregach czasowych oparte na sztucznej inteligencji w wybranych procesach złożonych. Zadanie to jest w nauce dość oczywiste i wypracowano wiele metod obficie cytowanych także w tej pracy. Cechą szczególną przedstawianej rozprawy jest nawiązywanie tymi zadaniami do bieżących społecznie ważnych problemów nurtujących cały świat w przypadku pandemii oraz Polskę, w przypadku inflacji. Zastosowane metody predykcyjne wyróżniają się w obu obszarach badań innowacyjnymi zmianami o autorskim charakterze.

Pandemia COVID-19 wywołana przez wirus SARS-CoV-2 przyniosła ludzkości więcej ofiar niż jakikolwiek ze znanych kataklizmów naturalnych, konfliktów zbrojnych oraz pandemii. Rozprzestrzeniający się po świecie wirus zbierał śmiertelne żniwo. Władze szczebla krajowego wprowadzały z różnym powodzeniem radykalne środki zapobiegawcze mające na celu ograniczenie zakażeń. Pandemia wywołała kryzys służby zdrowia w wielu krajach. Dotychczas stosowane metody prognozowania liczby zakażeń okazały się z różnych powodów niedokładne.

Celem rozprawy w pierwszym zadaniu dotyczącym pandemii było opracowanie metod i modeli statystycznych do prognozowania rozprzestrzeniania się wirusa SARS-CoV-2 na poziomie kraju z uwzględnieniem szczepień. Metody te pozwoliłyby na wczesną ewaluację podejmowanych polityk obostrzeń oraz przygotowanie służb do nadchodzących fal zakażeń. Dla potrzeb realizacji celu dokonano analizy literatury w zakresie wpływu pandemii na życie społeczne, gospodarcze oraz na środowisko naturalne. W drugim zadaniu dotyczącym inflacji zastosowano także oryginalną metodę predykcji i uzyskano zmniejszenie błędu prognozy w porównaniu z klasyczną metodą regresyjną. Dla obu obszarów badań dokonano przeglądu stosowanych metod prognozowania. Sformułowano cel i hipotezę badawczą pracy.

W części aplikacyjnej rozprawy zostały przedstawione 3 metody predykcji szeregów czasowych zakażeń SARS-CoV-2 oraz kryterium uwzględniające szczepienia – gęstość szczepień. Kolejne 2 metody z optymalizacją i korektą dotyczyły predykcji szeregów czasowych inflacji. W przypadku predykcji zakażeń metody bazowały na podobieństwach szeregów, w tym z uwzględnieniem szczepień. Dokładność prognozowania szeregów pandemicznych za pomocą proponowanych metod została poddana ewaluacji w zależności od horyzontu predykcji, dynamiki zakażeń, gęstości szczepień oraz położenia geograficznego badanego kraju. W metodzie prognozowania inflacji w szeregu czasowym opartym na notowaniach rok do roku wprowadzono autorską korektę prognoz w horyzoncie jednomiesięcznym.

Rozprawę kończą wnioski z przeprowadzonych w pracy rozważań i badań oraz wskazanie dalszych kierunków badań.

**Słowa kluczowe:** szeregi czasowe, predykcja, COVID-19, inflacja, metody statystyczne, procesy złożone, inteligencja obliczeniowa

## Abstract

The considerations in the doctoral dissertation concern the accuracy of statistical methods and models used for forecasting in time series based on artificial intelligence in selected complex processes. This task is quite obvious in science, and many methods have been developed, which are also abundantly cited in this work. A distinctive feature of the presented dissertation is that it links these tasks to current socially significant problems that concern the entire world in the case of the pandemic and Poland in the case of inflation. The applied predictive methods stand out in both research areas with innovative changes of an authorial nature.

The COVID-19 pandemic caused by the SARS-CoV-2 virus has brought more victims to humanity than any of the known natural disasters, war conflicts and pandemics. The virus spreading around the world was reaping a deadly toll. National authorities introduced, with varying degrees of success, radical preventive measures aimed at limiting infections. The pandemic caused a health service crisis in many countries. Previously used methods of forecasting the number of infections turned out to be inaccurate for various reasons.

The aim of the dissertation in the first task related to the pandemic was to develop methods and statistical models for forecasting the spread of the SARS-CoV-2 virus at the country level, taking into account vaccinations. These methods would allow for early evaluation of restrictive policies and preparation of services for upcoming waves of infections. For the purpose of achieving the goal, a literature review was conducted regarding the impact of the pandemic on social life, the economy, and the natural environment. In the second task concerning inflation, an original prediction method was also applied, resulting in a reduction of forecast error compared to the classical regression method. For both research areas, a review of the forecasting methods used was conducted. The goal and research hypothesis of the work were formulated.

In the application part of the dissertation, 3 methods of predicting time series of SARS-CoV-2 infections and a criterion taking into account vaccinations – vaccination density – were presented. The next 2 methods with optimization and correction concerned the prediction of time series of inflation. In the case of infection predictions, the methods were based on series similarities, including vaccinations. The effectiveness of pandemic series forecasting using the proposed methods was evaluated depending on the prediction horizon, infection dynamics, vaccination density, and the geographical location of the country studied. In the method of forecasting inflation in a time series based on year-to-year values, an authorial correction of forecasts in a one-month horizon was introduced.

The dissertation concludes with conclusions from the considerations and research conducted in the work and indicating further directions for research.

**Keywords:** time series, prediction, COVID-19, inflation, statistical methods, complex processes, computational intelligence

# 1 Wprowadzenie

Niniejsza rozprawa doktorska mieści się w obszarze informatyki (computer science) w ramach klasyfikacji ACM Computing Classification System („Computing Classification System”, <https://dl.acm.org/ccs> 2023), ponieważ skupia się na rozwoju i zastosowaniu zaawansowanych metod statystycznych i obliczeniowych do prognozowania zjawisk w szeregach czasowych (Rys. 1. ), które są kluczowymi komponentami informatyki. Praca ta wykorzystuje modelowanie matematyczne i uczenie maszynowe, które są podstawowymi narzędziami w informatyce stosowane do analizy danych i przewidywania trendów, co jest zasadnicze w kontekście big data i inteligencji obliczeniowej. Ponadto, opracowane metody mają szerokie zastosowania w praktycznych aspektach informatyki stosowanej, takich jak analiza rozprzestrzeniania się pandemii COVID-19 oraz prognozowanie zmian ekonomicznych, co bezpośrednio wpisuje się w rozwiązania systemów informacyjnych i ich wpływ na społeczeństwo i politykę.

Rozprawa doktorska jest podzielona na 4 główne części.

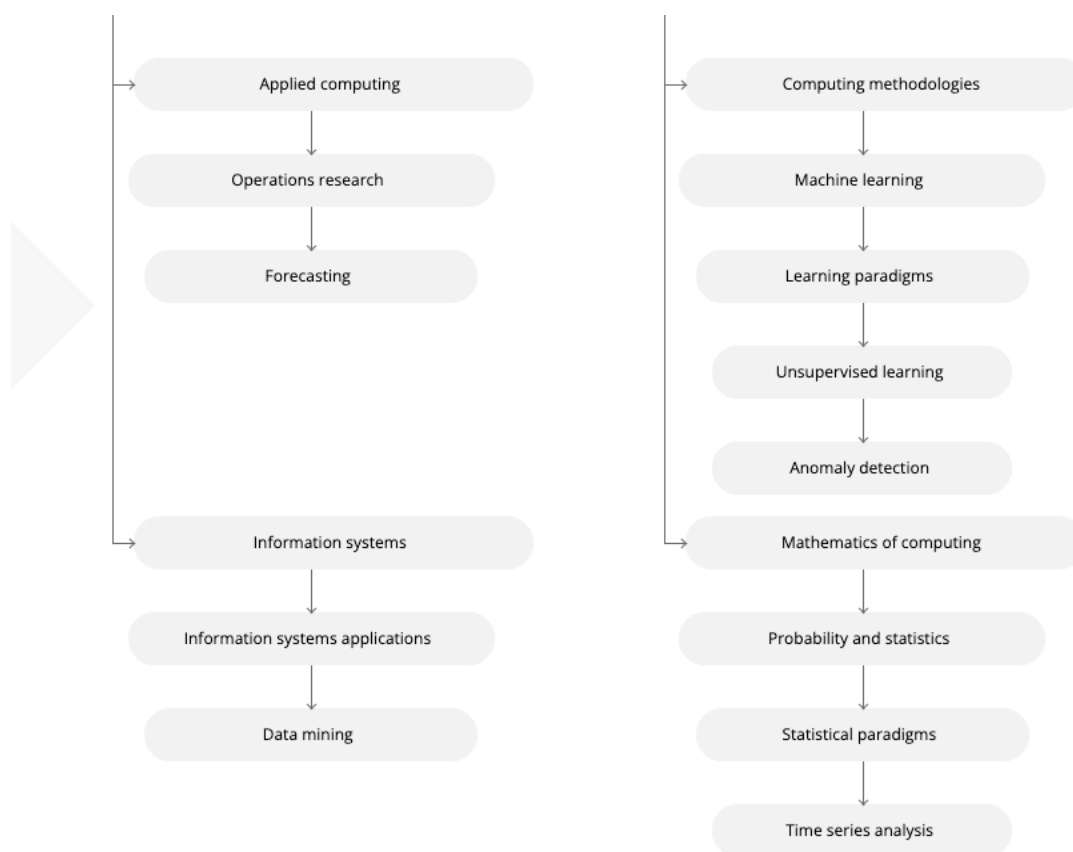
W pierwszej części przedstawiono cel pracy, sformułowano problem oraz postawiono hipotezę badawczą. Pandemia COVID-19 została przedstawiona jako jedna z największych odnotowanych w dziejach ludzkości. Przedstawiono wpływ pandemii na społeczeństwo oraz na środowisko naturalne. Przedstawiono ważną rolę, jaką odgrywają w tym trudnym czasie komunikacja społeczna, jakość danych oraz wdrażane obostrzenia. Następnie zapoznano się z działalnością krajowych zespołów eksperckich związanych z prognozowaniem rozprzestrzeniania się wirusa SARS-CoV-2. Kolejnym istotnym problemem społecznym jest inflacja, przedstawiono jej wpływ na społeczeństwo, gospodarkę oraz przedstawiono kształtowanie się inflacji w Polsce w ostatnich latach. W przypadku obu problemów społecznych wykonano przegląd literatury związanej z dotychczas stosowanymi metodami i modelami statystycznymi stosowanymi w predykcji ich szeregów czasowych oraz dokładnością ich prognoz.

W drugiej części zaproponowano 3 metody predykcji rozprzestrzeniania się wirusa SARS-CoV-2 oraz 2 metody z rozszerzeniami do predykcji inflacji w jednomiesięcznym horyzoncie predykcji. Przedstawiono ich model matematyczny, algorytm działania, sposób pozyskania i

przygotowania danych oraz sposób ewaluacji wyników. W przypadku pandemicznych szeregów czasowych zaproponowano kryterium gęstości szczepień mierzące poziom szczepień w krajach.

Trzecia część zawiera plan badań zdefiniowany w podejściu GQM oraz jego wykonanie. Ze względu na złożoność metod pandemicznych - wykonanie każdej z nich zostało udokumentowane w postaci licznych wykresów i tabel, w kolejnych uruchomieniach metod ograniczono się do przedstawienia najistotniejszych wniosków i obserwacji. Zbadano dokładność predykcji zaproponowanymi metodami w zależności od długości horyzontu predykcji, położenia geograficznego kraju, poziomu szczepień i dynamiki rozprzestrzeniania się wirusa w zdefiniowanych fazach pandemii. W przypadku inflacji przedstawiono zbiorcze wyniki metod dla poszczególnych krajów.

Ostania z części zawiera ewaluację osiągniętych celów oraz potwierdzenie hipotezy badawczej w formie konkluzji. Niniejszą rozprawę kończą propozycje dalszych kierunków badań.



Rys. 1. Miejsce niniejszej rozprawy doktorskiej w klasyfikacji ACM Computing Classification System. Źródło: „Computing Classification System”, <https://dl.acm.org/ccs> 2023.

## 1.1 Cel pracy

Celem pracy jest opracowanie i przetestowanie poprzez symulacje komputerowe narzędzia prognostycznego działającego w obszarze danych statystycznych pandemii COVID-19 oraz danych dotyczących inflacji. Narzędzie to będzie realizować zadania prognostyczne, umożliwiając wczytanie danych statystycznych dla dowolnego państwa i wykonanie prognoz wraz z ich wizualizacją. Dodatkowo, w kontekście inflacji, narzędzie ma na celu stworzenie precyzyjnych modeli predykcyjnych, które pozwolą na dokładną estymację przyszłych wskaźników inflacyjnych, mających istotne znaczenie dla kształtowania polityki monetarnej i ekonomicznej.

### 1.1.1 Motywacja

W obliczu globalnych wyzwań, jakie stawia przed nami XXI wiek, nauka staje przed zadaniem zrozumienia i przewidywania zjawisk o zasięgu globalnym, które mają bezpośredni wpływ na życie każdego z nas. Pandemia COVID-19 i jej konsekwencje, jak również dynamiczne zmiany w zakresie inflacji, są przykładami takich zjawisk. Wymagają one nie tylko głębokiej analizy, ale również opracowania narzędzi, które pozwolą na efektywne prognozowanie ich przyszłych kierunków. Tylko poprzez interdyscyplinarne podejście, łączące wiedzę z różnych dziedzin – od biologii i medycyny, przez ekonometrię, aż po informatykę – możemy mieć nadzieję na skuteczne zarządzanie kryzysami, które niosą ze sobą.

Ludzkość na przestrzeni lat zmagala się z licznymi chorobami pochodzenia wirusowego. W grudniu 2019 roku w Wuhan znajdującym się w chińskiej prowincji Hubei wykryto po raz pierwszy wirus SARS-CoV-2, który może przenosić się z człowieka na człowieka. Wirus ten powodował zakaźną chorobę układu oddechowego, znaną obecnie jako COVID-19. Choroba COVID-19 szybko rozprzestrzeniła się na cały świat [Wang et al. (2020)]. 13 marca 2020 roku WHO uznała oficjalnie COVID-19 jako pandemię. Choroba COVID-19 jest uznawana za największy kryzys globalny od czasów II wojny światowej. Porównywalna pod względem śmiertelności z pandemiemi globalnymi (hiszpanka, dżuma, HIV) i lokalnymi SARS, MERS i Ebola). Swoim zasięgiem choroba objęła 199 państw, powodując (do dnia 30 maja 2022 roku) 529 458 098 potwierdzonych zakażeń oraz 6 289 220 przypadków śmiertelnych („COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)”,

<https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6> (2022)). Jest unikalna ze względu na swój zasięg przestrzenny, szybki początek oraz złożoność jej konsekwencji.

Rozwój pandemii w krajach jest dotychczas nieprzewidywalny. Naukowcy z całego świata przeprowadzają badania w celu zrozumienia zależności rozwój choroby i rozprzestrzenianie się wirusa. Wpływ środowiska naturalnego oraz czynników demograficznych były jednym z popularnych kierunków badań. Malki et al. (2020) wykonał badania wpływu zmiennych pogodowych, gęstości zaludnienia oraz liczebności dostępnych miejsc w szpitalach na rozwój COVID-19. Tosepu et al. (2020) wskazał, że pogoda jest ważnym czynnikiem występowania COVID-19 lokalnie na przykładzie Jakarty. Średnia temperatury jest znacząco skorelowana z zakażeniami. Bashir et al. (2020) na przykładzie Nowego Jorku wskazuje znaczącą korelację pomiędzy średnią temperaturą, minimalną temperaturą i jakością powietrza a pandemią COVID-19. Kumar G. i Kumar R.R. (2020) wykazali korelację pomiędzy temperaturą minimalną, względną wilgotnością oraz ciśnieniem powietrza a COVID-19 dla Bombaju. Wykorzystali dane meteorologiczne do uczenia sztucznej sieci neuronowej i predykcji COVID-19. Cano-Pérez et al. (2020) wykazał odwrotną korelację pomiędzy położeniem - wysokością nad poziomem morza - a liczbą zakażeń oraz zgonów na przykładzie Kolumbii. Kolejną obserwacją był wpływ gęstości zaludnienia (wzmoczona aktywność ekonomiczna, większe środki transportu, skażenie powietrza, interakcje społeczne) na zwiększoną liczbę zakażeń. Jednocześnie na przykładzie Chin nie znaleziono dowodów na wpływ cieplejszej pogody na spadek liczby przypadków COVID-19 (Xie et al. (2020)).

Klasyczne modele predykcji rozwoju zakażeń SARS-CoV-2 w początkowej fazie pandemii nie odzwierciedlały specyfiki wirusa i dynamiki jego rozprzestrzeniania się, co wskazuje na konieczność opracowania nowych metod predykcyjnych. Niewystarczająca zdolność tych modeli do adaptacji do szybkich zmian i nowych wzorców zakażeń podkreśla potrzebę stworzenia narzędzi predykcyjnych, skupiających się bezpośrednio na szeregach czasowych liczby zakażeń.

Pandemia COVID-19, oprócz bezpośrednich skutków zdrowotnych, wywołała również szereg konsekwencji ekonomicznych, w tym niepokojące zjawisko inflacji. Przestoje w produkcji, zmiany w łańcuchach dostaw i zmieniające się zachowania konsumentów są tylko niektórymi z czynników, które wpłynęły na wzrost cen towarów i usług na całym świecie. Te dwa zjawiska –

zdrowotne i ekonomiczne – stały się nierozzerwalnie powiązane, tworząc nowe wyzwania dla prognozowania i planowania ekonomicznego.

Podobnie jak pandemia, inflacja jest zjawiskiem o globalnym zasięgu, które dotyka każdego obywatela, niezależnie od położenia geograficznego czy statusu ekonomicznego. Wzrost cen towarów i usług bezpośrednio wpływa na gospodarki krajów, wartość walut i budżety domowe, przez co staje się kluczowym wskaźnikiem ekonomicznym monitorowanym przez rządy i banki centralne. W ostatnich latach, w wyniku pandemii, obserwujemy nieoczekiwane skoki inflacyjne, które wymagają szczególnego podejścia analitycznego. Stabilność cen jest fundamentem zdrowej gospodarki, a jej wahania mogą prowadzić do poważnych zaburzeń finansowych i społecznych. Właściwe prognozowanie inflacji, zarówno w krótkiej, jak i długiej perspektywie, jest więc niezbędne dla utrzymania równowagi ekonomicznej i zapewnienia dobrobytu społeczeństwa

W ostatnich dziesięcioleciach metody prognozowania inflacji uległy znaczącym zmianom ze względu na dynamiczne i nieprzewidywalne zmiany w gospodarce światowej. W odpowiedzi na te wyzwania, istnieje pilna potrzeba opracowania nowych metod prognozowania inflacji, które będą bazować bezpośrednio na pojedynczym szeregu czasowym. Proste, jednowymiarowe podejście do szeregu czasowego inflacji umożliwia bezpośrednie skupienie się na ogólnych trendach inflacyjnych, bez komplikacji wynikających z analizy wielowymiarowych zestawów danych.

## 1.2 Pandemia i inflacja a procesy złożone.

W kontekście tematu niniejszej rozprawy doktorskiej istotne jest zrozumienie pojęcia „proces złożony” oraz uzasadnienie odniesienia do pandemicznych oraz inflacyjnych szeregów czasowych. Poniżej dokonano przeglądu literatury nawiązującej do pojęcia procesu złożonego, wymieniono cechy procesów i systemów złożonych oraz przedstawiono interpretację pojęcia zastosowaną w treści rozprawy.

Procesy złożone to interdyscyplinarne pojęcie będące przedmiotem rozważań naukowców. Pojęcie procesu złożonego jest powiązane z pojęciem systemu złożonego. **Proces złożony** to sekwencja zdarzeń lub działań, które mają złożoną strukturę i są często nieliniowe oraz trudne do przewidzenia ze względu na liczne zmienne i ich interakcje. Procesy złożone mogą zachodzić w ramach systemów złożonych i są odpowiedzialne za ich ewolucję, adaptację oraz za emergentne

zachowania. **System złożony** odnosi się do struktury lub zbioru elementów, które są powiązane i wzajemnie na siebie oddziałują. Systemy takie charakteryzują się dynamiką i zazwyczaj wykazują zachowania emergentne, czyli takie, które nie są oczywiste na podstawie właściwości pojedynczych składników systemu.

W literaturze brak jest jednoznacznej definicji określającej proces (lub system) złożony. Można jednak zidentyfikować cechy wspólne pomiędzy definicjami. Za prekursora określenia systemu złożonego uważany jest Simon (1962), który zdefiniował system złożony jako taki, który składa się dużej liczby części, które wchodzi z sobą w interakcje w nietrywialny sposób. Autor wskazuje, że w takich systemach całość jest czymś więcej niż suma jej części, przy czym nie odnosi się tu do sensu metafizycznego. Wskazuje na sens praktyczny, w którym zrozumienie właściwości części i praw rządzących ich interakcją nie prowadzi w prosty sposób do zrozumienia właściwości całego systemu.

Procesy złożone są definiowane współcześnie w kontekście systemów dynamiki nieliniowej, teorii chaosu, teorii systemów oraz innych dziedzin naukowych. Istnieje kilka kluczowych teorii i podejść, które umożliwiają zrozumienie ich charakterystyk i dynamiki.

1. Zgodnie z **teorią systemów złożonych**, procesy te charakteryzują się interakcjami między wieloma elementami, co prowadzi do emergencji, czyli pojawiania się nowych właściwości, które nie są widoczne na poziomie poszczególnych składników (Miller & Page (2007)), Bar-Yam (1997). Przykładem mogą być globalne trendy ekonomiczne, takie jak inflacja, które są wynikiem złożonych interakcji między różnymi gospodarkami, politykami monetarnymi i zachowaniami rynkowymi.
2. Z kolei w kontekście **dynamiki nieliniowej i teorii chaosu**, procesy złożone są opisywane jako systemy, w których niewielkie zmiany w warunkach początkowych mogą prowadzić do znacząco różnych wyników, utrudniając długoterminowe prognozy (Strogatz (2014)), (Gleick (1987)). Jest to szczególnie istotne w kontekście rozprzestrzeniania się chorób, gdzie początkowe wzorce zakażeń mogą prowadzić do bardzo różnych trajektorii epidemii.
3. Dodatkowo, **teoria sieci złożonych** oferuje perspektywę analizy procesów złożonych jako sieci składających się z węzłów i połączeń, gdzie struktura sieci ma znaczący wpływ na dynamikę i zachowanie całego systemu (Newman (2010)), (Barabási (2016)). W kontekście pandemii, sieć taka może uwzględniać połączenia transportowe, interakcje

społeczne i mobilność ludności, które wszystkie wpływają na dynamikę rozprzestrzeniania się choroby.

Sayama (2015) określa systemy złożone jako sieci składające się z wielu komponentów, które oddziałują ze sobą, zazwyczaj w sposób nieliniowy. Systemy złożone mogą powstawać i ewoluować przez samoorganizację, tak że nie są ani całkowicie regularne, ani całkowicie przypadkowe, co umożliwia rozwój zachowań emergentnych na makroskopowej skali.

Ladyman et al. (2013) wskazują na zbiór cech jakimi odpowiada system złożony. Autorzy potwierdzają, że system złożony jest pojęciem interdyscyplinarnym, którego definicja jest trudna do ustalenia i różnie interpretowana w różnych dziedzinach nauki. Podają podstawowe cechy systemów złożonych. Podkreślają zrozumienie, że te cechy nie są ani wystarczające, ani wymagane do klasyfikacji systemu jako złożonego. Są to raczej ogólne wytyczne pomagające zrozumieć charakter i działanie takich systemów.

1. **Nielinearność.** Zachowanie systemu nie jest proporcjonalne do wpływających na niego bodźców.
2. **Sprzężenie zwrotne.** Systemy złożone często zawierają pętle sprzężeń zwrotnych, które mogą wzmacniać lub tłumić ich działanie.
3. **Porządek spontaniczny.** Systemy złożone mogą samorzutnie organizować się w bardziej złożone struktury.
4. **Odporność i brak centralnego sterowani.** Systemy złożone często są odporne na zakłócenia i nie posiadają centralnego elementu kontrolnego.
5. **Emergencja.** Nowe właściwości mogą pojawiać się na wyższych poziomach organizacji systemu, które nie są widoczne na niższych poziomach.
6. **Hierarchiczna organizacja.** Systemy złożone często wykazują wielopoziomową strukturę organizacyjną.
7. **Duża liczba elementów.** Systemy złożone składają się z wielu wzajemnie oddziałujących na siebie komponentów.

Bobkowska (2014) określiła cechy odpowiadające obiektowi (systemowi, procesowi) złożonemu:

1. **Działanie systemu ma charakter stochastyczny.** Oznacza to, że system wykazuje pewną losowość w swoim zachowaniu i nie jest całkowicie przewidywalny. Procesy stochastyczne są zasadniczymi składnikami w modelowaniu systemów złożonych,

ponieważ uwzględniają przypadkowość i niepewność wynikającą z wielu czynników, takich jak zewnętrzne wpływy, wewnętrzna zmienność elementów systemu, czy ograniczenia informacyjne.

2. **Użyteczny system działa w taki sposób, że jego działania skierowane są na stopniowe zmniejszenie entropii informacyjnej** (właściwość tzw. nieentropijności).
3. **Właściwości systemu nie są pochodnymi właściwości jego podsystemów.**
4. **System jest emergentny.** System emergentny to rodzaj systemu złożonego, w którym pojawia się emergencja, czyli zjawisko, w którym całość systemu wykazuje nowe właściwości, które nie są widoczne ani przewidywalne tylko na podstawie analizy pojedynczych elementów składowych tego systemu. Takie systemy charakteryzują się złożonymi wzajemnymi interakcjami między swoimi składnikami, prowadząc do powstania nowych wzorców, struktur, zachowań lub funkcji, które nie mogą być wyjaśnione wyłącznie przez właściwości poszczególnych elementów.
5. Cele użytecznego systemu polegają na znalezieniu zadowalających, ale niekoniecznie najlepszych rozwiązań spośród wszystkich hipotetycznie możliwych.

Reasumując, w niemiejszej rozprawie proces złożony zostaje zdefiniowany jako sieć wzajemnie połączonych struktur, których skomplikowane wzajemne oddziaływania wykraczają poza możliwość prostego opisu za pomocą konwencjonalnych równań matematycznych. Biorąc pod uwagę różnorodność i zawłość tych struktur, a także fakt, że ich relacje często pozostają niewyjaśnione lub są poza zasięgiem identyfikacji, stosujemy podejście do tworzenia modeli oparte na symulacjach. Te symulacje są konstruowane i przeprowadzane z wykorzystaniem szeregów czasowych, które dostarczają danych wynikowych, będących odzwierciedleniem dynamiki badanego systemu. Szeregi czasowe, jako narzędzie do rejestrowania zmian w czasie, są kluczowe dla modelowania ewolucji i adaptacji badanych procesów. Istotne jest precyzyjne dostosowanie danych w taki sposób, by ich zbieranie, przetwarzanie i analiza odbywały się w sposób ujednolicony, co zapewnia integralność i porównywalność informacji wykorzystywanych do odwzorowania złożonych dynamik procesów w predykcji szeregów czasowych.

Prognozowanie infekcji SARS-CoV-2, jak i prognozowanie inflacji można uznać za złożone systemy ze względu na ich nieliniowe interakcje, naturę dynamiczną, wrażliwość na warunki

początkowe i zachowania emergentne. Te cechy sprawiają, że dokładne prognozowanie jest wyzwaniem i wymaga zastosowania zaawansowanych technik modelowania z wykorzystaniem uczenia maszynowego i sztucznej inteligencji, aby uchwycić podstawową dynamikę tych złożonych systemów.

Prognozowanie w szeregach czasowych **zakażeń** jest procesem złożonym ze względu na:

1. **Nieliniowe interakcje.** Na rozprzestrzenianie się wirusa ma wpływ mnóstwo czynników, w tym zachowania społeczne, polityki rządowe, warunki środowiskowe i biologiczne cechy wirusa. Te czynniki oddziałują na siebie w nieliniowy sposób, co czyni prognozowanie wskaźników infekcji wysoce złożonym.
2. **Dynamiczna natura.** Pandemia jest dynamiczna, z wskaźnikami infekcji zmieniającymi się w zależności od różnych czynników, takich jak tempo mutacji wirusa, wskaźniki szczepień i możliwości służby zdrowia.
3. **Wrażliwość na warunki początkowe.** Małe zmiany w przenośności wirusa lub zachowaniu publicznym mogą prowadzić do znacznie różnych wyników w zakresie wskaźników infekcji.
4. **Zachowanie emergentne.** Pandemia wykazuje zachowanie emergentne, takie jak rozwój nowych szczepów wirusa czy wpływ środków dystansu społecznego, które nie są łatwo przewidywalne wyłącznie na podstawie poszczególnych składników.

Z kolei prognozowanie w szeregach czasowych **inflacji** jest procesem złożonym ze względu na:

1. **Liczne oddziałujące zmienne.** Inflacja jest wynikiem szerokiego zakresu czynników ekonomicznych, np. stopy procentowe, wskaźniki zatrudnienia, koszty produkcji i nawyki konsumpcyjne konsumentów. Te zmienne oddziałują na siebie w złożony i często nieliniowy sposób.
2. **Adaptacyjność.** Gospodarka jest adaptacyjna, reagując na zmiany polityki, trendy rynkowe i szoki zewnętrzne (jak pandemia, zmiany polityczne).
3. **Dynamika i nieprzewidywalność.** Wskaźniki ekonomiczne, takie jak CPI, są dynamiczne i mogą być wysoce zmienne, reagując zarówno na wydarzenia gospodarcze krajowe, jak i międzynarodowe.

4. **Właściwości emergentne.** Systemy ekonomiczne wykazują właściwości emergentne, takie jak inflacja czy recesja, które wynikają ze złożonej interakcji różnych działalności gospodarczych.

Podobnie jak wskazywany przez Bobkowska (2014) proces adaptacyjnej budowy modeli procesów złożonych w nierniejszej rozprawie obejmuje:

1. Identyfikacja i zbieranie danych z dostępnych źródeł.
2. Analizę zebranych informacji, oczyszczanie danych i eliminację danych nieistotnych.
3. Pozyskanie wiedzy o obiekcie badań (przegląd literatury).
4. Tworzenie modeli w drodze symulacji ewolucyjnej lub samoorganizacji modeli.
5. Predykcja zachowania obiektu w zdefiniowanym horyzoncie predykcji.

### 1.3 Społeczne znaczenie inflacji i jej prognoz

Inflacja jest ważnym zjawiskiem gospodarczym, które ma wpływ na wiele aspektów życia społecznego i gospodarczego. Poniżej przedstawiono społeczne znaczenie inflacji, w tym jej negatywne skutki prowadzące do spadku standardu życia, zniechęcenia do oszczędzania i inwestycji oraz pogorszenia sytuacji finansowej osób o niskich dochodach. Wskazano wpływ inflacji oraz jej prognoz na decyzje o wymiarze mikroekonomicznym i makroekonomicznym.

Inflacja bezpośrednio wpływa na siłę nabywczą konsumentów, zmniejszając realną wartość pieniądza. Wzrost cen towarów i usług oznacza, że za tę samą kwotę można nabyć mniej dóbr. To zjawisko szczególnie dotyka osoby o stałych dochodach, które nie są dostosowywane do tempa inflacji (Coibion et al. (2020)). Wysoka inflacja może prowadzić do spadku standardu życia, gdyż konsumenci są zmuszeni do ograniczania wydatków na niektóre dobra i usługi. Długotrwała wysoka inflacja może również prowadzić do zmniejszenia zaufania konsumentów i spadku wydatków konsumpcyjnych, co negatywnie wpływa na gospodarkę (D'Acunto et al. (2021)).

Inflacja ma znaczący wpływ na oszczędności, ponieważ ich realna wartość maleje wraz ze wzrostem inflacji. Dla osób oszczędzających może to oznaczać utratę części ich majątku. Wysoka inflacja może zniechęcać do oszczędzania, ponieważ stopa zwrotu z oszczędności jest niższa niż stopa inflacji, co w efekcie może przełożyć się na zmniejszenie inwestycji w gospodarce i spowolnienie wzrostu gospodarczego.

Emeryci i renciści są grupami społecznymi narażonymi na poważne konsekwencje wzrostu inflacji, które często są ustalane na stałym poziomie i nie są dostosowywane do bieżącej inflacji. Wzrost inflacji może prowadzić do spadku realnej wartości tych świadczeń, co wpływa na dobrobyt osób starszych (Sarbiński (2016)). W niektórych przypadkach, systemy emerytalne są indeksowane do inflacji, ale często z opóźnieniem, co nadal może prowadzić do spadku siły nabywczej emerytur.

Inflacja wpływa na rynek pracy. W krótkim okresie, niska inflacja może być postrzegana jako korzystna dla zatrudnienia, ponieważ firmy mogą zwiększać produkcję bez znaczącego wzrostu kosztów. W dłuższej perspektywie, wysoka inflacja wywołuje ryzyko inwestycyjne zniechęcające do zatrudnienia i wzrost bezrobocia.

Wysoka inflacja może obniżać realną wartość długu publicznego i prywatnego, co jest korzystne dla dłużników, ale niekorzystne dla wierzycieli. Stopy procentowe zwykle rosną w środowisku wysokiej inflacji, co zwiększa koszty pożyczek i może prowadzić do problemów z zadłużeniem zarówno w sektorze publicznym, jak i prywatnym.

Prognozowanie inflacji jest istotne z wielu powodów, zarówno dla przedsiębiorców, rządu jak i gospodarstw domowych. Precyzyjne prognozy inflacji pozwalają rządowi i bankom centralnym na lepsze planowanie polityki monetarnej i fiskalnej (Próchnicki (2011)).

Ciżkowicz, P., & Rzońca, A. (2011) odnoszą się do podejścia rządów do zarządzania inflacją w sytuacji szybko rosnących cen. Wskazują, że rządy często ogłaszają nierealistyczne plany ograniczenia wzrostu cen i obiecują spadek inflacji, ale nie podejmują wystarczających działań zaostrzających politykę pieniężną, które mogłyby ten spadek wywołać. Ponadto zauważono, że trudno ludziom uwierzyć w zapewnienia rządu, który w przeszłości nie potrafił utrzymać inflacji pod kontrolą. Rząd, który nie zdołał kontrolować inflacji, jest postrzegany jako mniej wiarygodny niż ten, który taką kontrolę zachowa

Przedsiębiorstwa i inwestorzy wykorzystują prognozy inflacji do planowania inwestycji, oceny ryzyka, wydatków i cen. Dzięki temu mogą oni minimalizować negatywny wpływ inflacji. W sytuacji niepewności związanej z inflacją przedsiębiorcy często wstrzymują inwestycje. Jest to spowodowane tym, że decyzje inwestycyjne są długoterminowe i często trudno jest się z nich wycofać bez dodatkowych kosztów (Ciżkowicz, P., & Rzońca, A. (2011)).

Dokładne prognozy pomagają w utrzymaniu stabilności cen, co jest kluczowe dla kondycji gospodarczej kraju. Skuteczna polityka pieniężna, w tym zarządzanie podażą pieniądza, stopami

procentowymi i kursami walutowymi, jest kluczowa dla stabilności finansowej. Kontrola inflacji, wspieranie inwestycji oraz tworzenie stabilnego środowiska gospodarczego są niezbędne do utrzymania zdrowej równowagi ekonomicznej i finansowej (Chaudhry et al. (2021)).

Prognozy inflacji mogą pomóc konsumentom i inwestorom w podejmowaniu decyzji o sposobie ochrony swoich oszczędności przed utratą wartości. Istnieje długoterminowa równowaga między oszczędnościami, stopą inflacji i wzrostem gospodarczym. Zmiany w jednej zmiennej mogą mieć wpływ na pozostałe w długim okresie (Er et al. (2014)). Nowaczewska (2023) wskazuje na przykładzie Polski, że gospodarstwa domowe mogą chronić oszczędności poprzez inwestycje np. w złoto, nieruchomości oraz powszechnie dostępne złoto oraz akcje.

Prognozy inflacji mogą być wykorzystywane do planowania oraz negocjowania płac i wynagrodzeń. Jeśli inflacja jest wysoka, pracownicy mogą wymagać wyższego wzrostu płac, aby utrzymać swoją siłę nabywczą. Szczepanik (2011) stwierdził, że w Polsce wzrost wynagrodzeń nominalnych słabo, ale pozytywnie koreluje z inflacją, wpływając na nią z kwartalnym opóźnieniem, co wskazuje na wpływ popytu na inflację. Nie zaobserwował istotnego wpływu inflacji na poziom wynagrodzeń, co oznacza, że wzrost płac bardziej wpływa na inflację niż odwrotnie.

## 1.4 Wpływ wybranych pandemii na społeczeństwo

Pandemie miały znaczący wpływ na społeczeństwa. Piret i Boivin (2021) oraz LePan (2020) LePan dokonali przeglądu znanych pandemii w historii ze wskazaniem wektorów ich rozprzestrzeniania. Większość znanych pandemii bazuje na wektorze odzwierzęcym (np. pchły, gryzonie, nietoperze, ptaki, świnie), ale część z nich jest spowodowana działalnością człowieka i jego wpływem na środowisko naturalne, np. epidemie dżumy spowodowane skażeniem wody. Spośród znanych pandemii i epidemii wyróżniającymi się były:

- Dżuma Justyniana
- Czarna śmierć zwana także dżumą
- Pandemie cholery odnotowywane w latach 1817 – 1923 oraz trwająca ostatnia siódma pandemia od 1961 roku do dnia dzisiejszego.
- Grypa „hiszpanka”
- AIDS

- COVID-19

Niewątpliwie wszystkie z powyżej wymienionych epidemii i pandemii niosły ze sobą wiele ofiar. Kluczowym dla ludzkości było zrozumienie ich powstawania, leczenie, opracowanie środków zapobiegawczych oraz wpływ, jaki wywarły na daną populację i środowisko naturalne. Niewątpliwie istotne było gromadzenie danych oraz ich analiza, które z biegiem wieków nabierało znaczenia. Na podstawie obserwacji i analizy danych można było znaleźć wytłumaczenie na stawiane pytania, z którymi ówczesna „medycyna” nie mogła sobie poradzić. Poniżej przedstawiono pandemię z perspektywy danych, w tym: śmiertelności, czynników wpływających na rozwój pandemii oraz wpływu na ekonomię. Wskazano komunikację - dezinformację społeczną w czasie konfliktu zbrojnego przypadającego na czas rozwoju grypy „hiszpanki”. Przedstawiono również niezaprzeczalny przykład rozwiązania problemu medycznego za pomocą analizy statystycznej opartej na danych było określenie źródła przenoszenia cholery.

#### 1.4.1 Dżuma Justyniana

Pierwszą wskazaną pandemią była Dżuma Justyniana szacowana na lata 541-543 i spowodowała około 100 milionów ofiar w Europie, natomiast w samym Konstantynopolu dzienna liczba zgonów spowodowanych skutkami zakażenia wynosiła 5-10 tysięcy osób [Zietz i Dunkelberg (2004)]. Wirus był przenoszony przez pchły zarażone od dzikich gryzoni. Duży wpływ na rozwój pandemii miały obserwowane w tym okresie zmiany klimatyczne wpływające na braki zbóż w spichlerzach oraz populację gryzoni [Wagner et al. (2014)].

#### 1.4.2 Czarna śmierć

Kolejną znaną pandemią z czasów średniowiecza była dżuma, zwana szerzej w krajach niemiecko- oraz angielskojęzycznych jako czarna śmierć. Varlık, N. (2021) badał pochodzenie ponurego określenia dżumy. Nazwa ta została ujęta po raz pierwszy w 1832 r. w pracy na temat czarnej śmierci niemieckiego lekarza i historyka medycyny Justusa Friedricha Carla Heckera. Hacker tworząc termin „czarna śmierć” chciał nadać emocjonalny ton epidemiologii historycznej jako epidemiologii gotyckiej.

Wpływ tej choroby na ekonomię badał Routt (2008). Choroba ta została początkowo wykryta we Włoszech w 1347 r. i przetoczyła się przez Europę aż do jej wygaśnięcia w 1353 roku

na terenach Rosji. Liczba śmiertelnych przypadków czarnej śmierci plasuje się pomiędzy 17 a 28 milionów. Nieliczne zachowane źródła przedstawiają apokaliptyczne wizje przebiegu tej pandemii. Śmiertelność była zróżnicowana ze względu na stan społeczny oraz położenie geograficzne. Przykładowo śmiertelność wśród Angielskiej arystokracji wynosiła 5-23,6%, natomiast na kontynencie 60% wśród duchownych Barcelony (Hiszpania) oraz 70% wśród duchownych z Bremen (Niemcy). Czarna śmierć miała istotny wpływ na średniowieczne rolnictwo. Zmniejszenie populacji umożliwiło zmniejszenie skali produkcji, zwiększenie dostępności ziemi, skoncentrowanie produkcji na żyznych glebach oraz odłogowanie. W handlu i przemyśle zwiększyła się konkurencyjność, co sprzyjało rozwojowi technologicznemu: mechanizacja produkcji odzieży, produkcja broni palnej, podstawy optyki oraz druku [Harris i Moss (2020)].

#### 1.4.3 Cholera – 7 pandemii

Cholera została wykryta w Azji w 1817 roku w Indiach. Analizy pandemii cholery dokonali Poiret i Boivin (2021). Globalizacja i rozwój technologiczny, a zwłaszcza transportu morskiego i kolejowego, wpłynął na szybkie rozprzestrzenianie się choroby. Kolejne pandemie cholery pojawiały się w różnych częściach świata i były powodowane przez odmienne bakterie. Rzetelne zbieranie i analiza danych umożliwiły ustalenie źródła zakażeń w trakcie drugiej pandemii po jej wybuchu w Londynie w 1854 roku. Lekarz John Snow po raz pierwszy wykorzystał epidemiologiczne metody do śledzenia czasu i rozprzestrzeniania się pandemii w mieście. Powiązał ogniska zakażeń z lokalizacją publicznych pomp wody, co umożliwiło mu wywnioskowanie, że to woda jest źródłem zakażeń. Zaproponował środki zapobiegawcze rozprzestrzenianiu się choroby w postaci usunięcia uchwyty do pomp w miejscach występowania ognisk zakażeń [Smith (2002)]. Bakteria cholery została wyizolowana dopiero w 1884 roku przez Roberta Kocha, natomiast toksyna odpowiedzialna za rozprzestrzenianie się choroby została odkryta w 1959 roku. Robert Koch zrozumiał wpływ zanieczyszczeń wody na rozprzestrzenianie się bakterii. Obecność cholery wiązana jest ze złymi warunkami bytowymi związanymi z ograniczonym dostępem do wody pitnej, przeludnieniem i brakiem wydajnej kanalizacji sanitarnej. Częstym powodem występowania okoliczności sprzyjających wybuchowi zakażeń cholery są uszkodzenia infrastruktury spowodowane klęskami żywiołowymi (np.

trzęsienia ziemi, powodzie) lub działalnością człowieka (niezrównoważony wpływ na środowisko, konflikty zbrojne). Wartą wspomnienia jest siódma pandemia cholery, wywołująca lokalne epidemie np. w Zimbabwie (2008), Haiti (2010) i Jemenie (2016). WHO odnotowało 923 037 zakażeń i 1911 zgonów w 31 krajach. Ryzyku rozprzestrzeniania się choroby sprzyja ocieplający się klimat.

#### 1.4.4 Grypa „hiszpanka”

Kolejną znaną pandemią o zasięgu globalnym była „hiszpanka” - wykryta po raz pierwszy w 1918 roku. Nazwa tej pandemii pochodzi od kraju, który otwarcie komunikował dane o tej chorobie. Pozostałe kraje, często uwikłane w konflikt wojenny, unikała informowania o liczbach związanych z tą chorobą, a nawet świadomie prowadziły dezinformację. Często zwykłe manipulowały swoimi społeczeństwami poprzez brak podawania rzetelnych danych, co również można zaobserwować w pandemii COVID-19 [Adiguzel et al. (2020)]. Hiszpanka spowodowała śmierć 50 milionów ludzi w przeciągu zaledwie 2 lat (Taubenberger i Morens (2006)).

#### 1.4.5 HIV i AIDS

Współcześnie jednym z najbardziej śmiertelnych wirusów jest HIV i wywoływana nim choroba AIDS. W 2020 roku na świecie było 37,7 miliona osób zakażonych wirusem HIV (dane notowane od 1980 roku), natomiast 36,3 miliona osób zmarło na AIDS od początku epidemii („Global HIV & AIDS statistics — Fact sheet” <https://www.unaids.org/en/resources/fact-sheet>).

### 1.5 Wpływ pandemii COVID-19 na środowisko naturalne

Efekty COVID-19 są rozpatrywane zwykle negatywnie z punktu widzenia wpływu na populację ludzką. Niewątpliwie pandemia pochłonęła wiele ludzkich żyć, ich śmierć była tragiczna z perspektywy rodzin. Pomimo niezaprzeczalnego tragizmu pandemii w ludzkim wymiarze, poniżej przedstawiono pozytywne aspekty pandemii. Jednym z nich w przeważającej mierze pozytywny jest wpływ pandemii na środowisko naturalne, w tym wpływ na jakość powietrza i zmniejszenie emisji hałasu. Pandemia przyspieszyła wprowadzenie transformacji cyfrowej, która przełożyła się na rozwój usług świadczonych drogą elektroniczną, upowszechnienie pracy, nauki i medycyny zdalnej oraz komunikacji na odległość. Niewątpliwie

do rozwoju transformacji cyfrowej przyczyniły się wprowadzane restrykcje na podstawie polityk rządowych. Powszechnie stosowanymi restrykcjami były kwarantanna (przedstawiono historyczne pochodzenie tego ograniczenia), izolacja, ograniczenie mobilności. Wskazano wybrane metody na pozyskiwanie danych związanych z ograniczeniami oraz monitorowania społeczeństwa. Dane liczbowe dotyczące tych ograniczeń są powszechnie stosowane do symulacji i predykcji rozprzestrzeniania się wirusa oraz skuteczności wprowadzanych środków zapobiegawczych. Podkreślono również kwestie dotyczące komunikacji społecznej, jakości i spójności danych, co miało wpływ na precyzję prognozowania rozwoju pandemii i efektywność implementowanych strategii zwalczania pandemii.

### 1.5.1 Środowisko naturalne

Największy wpływ na środowisko naturalne wywierały konsekwencje zamykania lub znacznego ograniczania sektorów ekonomicznych jak przemysł ciężki, transport i hotelarstwo Cheval et al. (2020). COVID-19 wpłynął na społeczeństwo w sposób radykalny, zmieniając styl życia ludzi i społeczeństwo, przyspieszając zmiany, które i tak najprawdopodobniej by się dokonały w dłuższej perspektywie czasowej („The recovery from the COVID-19 crisis must lead to a different economy”, <https://www.un.org/en/un-coronavirus-communications-team/launch-report-socio-economic-impacts-covid-19>).

Cheval et al. (2020) dokonał analizy wpływu COVID-19 na środowisko. Podzielił on wpływ ze względu na:

- perspektywę czasową:
  - natychmiastowy (dni i tygodnie),
  - krótkoterminowy (miesiące)
  - długo terminowy (lata)
- zasięg:
  - lokalny,
  - regionalny,
  - globalny,
- oraz wpływ:
  - negatywny, pozytywny i neutralny.

Tab. 1. Matryca przedstawiająca zaobserwowane oraz potencjalne wpływy COVID-19 na zmiany klimatu oraz środowiska. Czerwonym kolorem wskazano negatywne wpływy, zielonym pozytywne a szarym neutralne [Cheval et al. (2020)].

	Days to weeks	Months	Years
Global	Reduction of observational data and monitoring capacity	Societal and economic changes with environmental impacts	Environmental and climate change policies
		Environmental research for epidemiological applications	
Regional	Decreased environmental pollution due to reduced traffic	Enhanced systems for environmental monitoring	
	Shoreline pollution with sanitary disposal	Regional climate policy	Improved regional climate
		Improved adaptation plans	Improvements of citizen's wellness and health protection
			Enhanced planning for disaster risk reduction
Local		Ecological restoration	Changes in water consumption share
	Decreased noise pollution		
	Increased water and soil pollution due to waste disposal	Improved adaptation plans for cities	
		Improved air quality in urban areas	Changes in societal behaviour
	Immediate	Short-term	Long-term

W dokonanej analizie przewyższa wpływ pozytywny. Na terenach zurbanizowanych zaobserwowano znacznie poprawioną jakość powietrza, zmniejszony poziom hałasu oraz poziom skażeń ze względu na ograniczenia w ruchu ulicznym. W samym Delhi, stolicy Indii, odnotowano spadek hałasu rzędu 40-50% (Somani et al. (2020)). Poziom hałasu zmniejszył się ze 100 dB do 40 dB w porze dziennej i 30 dB w porze nocnej, co umożliwiło mieszkańcom słyszenie ćwierkania ptaków zwykle mieszczących się pomiędzy 40 a 50 dB (Rume i Islam (2020)). W samych Niemczech ograniczenia związane z przemieszczaniem się ograniczyły ruch lotniczy o 90%, uliczny o ponad 50% i kolejowy poniżej 25% (Sims (2020)).

Na poprawienie jakości powietrza wpływ również miało czasowe zamknięcie fabryk. W Chinach w czasie najdłuższego okresu zamknięcia fabryk odnotowano zmniejszenie emisji CO<sub>2</sub> o 25% („Economic slowdown as a result of COVID is no substitute for Climate Action”, <https://public.wmo.int/en/media/news/economic-slowdown-result-of-covid-no-substitute->

climate-action (2020)). Natomiast ogólna emisja NO<sub>2</sub> zmalała w Chinach o 20,2% od stycznia do maja 2020 w porównaniu do tego samego okresu w poprzedzającym roku. W USA poziom emisji NO<sub>2</sub> zmalał o 23,1% w lutym oraz 14,3% w marcu 2020 roku w porównaniu do tego samego okresu w roku poprzednim. Podobne spadki emisyjności NO<sub>2</sub> odnotowano w Wielkiej Brytanii, Francji, Niemczech i Włoszech (Liu et al. (2020)).

Wśród krótko i długoterminowych skutków pozytywnych badania środowiskowe dla zastosowań epidemiologicznych, rozbudowę systemów monitorowania środowiskowego oraz poprawione plany adaptacyjne dla miast. Kolejnymi pozytywnymi aspektami jest odtworzenie ekologiczne, ulepszenie ochrony zdrowia oraz regionalnych klimatów.

Negatywne oddziaływanie COVID-19 zostało wykazane tylko w natychmiastowej perspektywie i objawia się poprzez zwiększenie zanieczyszczenia wybrzeży, gleb i wody odpadami w tym sanitarnymi oraz zmniejszeniem możliwości monitorowania danych obserwacyjnych. Takimi danymi obserwacyjnymi są np. dane pozyskiwane z samolotowych obserwacji meteorologicznych. ECMWF odnotowała spadek o 65% samolotowych raportów meteorologicznych pomiędzy 3 a 23 marca 2020 roku („Drop in aircraft observations could have impact on weather forecasts”, <https://www.ecmwf.int/en/about/media-centre/news/2020/drop-aircraft-observations-could-have-impact-weather-forecasts> (2020)). WMO w kwietniu 2020 roku wyraziło niepokój o wpływie COVID-19 na ilościowe i jakościowe obserwacje i prognozy pogodowe, jak i również monitorowanie klimatu i atmosfery („Impacts on the Global Observing System”, <https://public.wmo.int/en/resources/coronavirus-covid-19/impacts-global-observing-system> (2020)). Z drugiej strony, radykalne zmniejszenie ludzkiej aktywności umożliwiło znaczne obniżenie szumu sejsmicznego umożliwiając dokładniejsze wykrywanie fal sejsmicznych i powtórnych trzęsień ziemi (Gibney (2020)).

Skutkami wskazanymi jako neutralne są: zmiany w zużyciu wody i zagospodarowaniu świateł (Paleologos et al. (2020)), w zachowaniach społecznych oraz zmianach polityk środowiskowych. Bashir et al. (2020) wskazuje, że twórcy regulacji i polityk powinni wprowadzić znaczne zmiany w regulacjach środowiskowych w celu ochrony dokonanych zmian w środowisku. Wskazuje na regulacje w celu zapewnienia ochrony przed COVID-19 pracowników służby zdrowia, osób w podeszłym wieku i dzieci jako grup zawodowych i wiekowych najbardziej narażonych na infekcje.

## 1.5.2 Rozwój technologiczny

Wybuch pandemii COVID-19 zainicjował w społeczeństwie nagły, dramatyczny i ekstensywny proces transformacji cyfrowej. Wprowadzenie ograniczeń i polityk walki z COVID-19 wywołało szok społeczny i paraliż instytucji. Jednymi z najbardziej dotkniętych przez następstwa COVID-19 były sektory usług publicznych, edukacja i szkolnictwo wyższe, sektor finansowy, bankowy, mobilność i usługi. Odpowiedzią na te wyzwania często była transformacja cyfrowa.

Wyrazistym przykładem jest edukacja. Uczniowie, nauczyciele oraz rodzice (zwłaszcza tych dzieci w wieku wczesnoszkolnym) musieli nagle nabyć nowych kompetencji i zdobyć zasoby w celu zintegrowania cyfrowych narzędzi w procesie nauki i uczenia się (Ivari et al. (2020)). Zbarachewicz (2020) badała problemy związane z wprowadzeniem transformacji cyfrowej w czasie pandemii COVID-19 na uczelniach wyższych. Uczelnie wyższe zmuszone zostały do wprowadzenia nowych rozwiązań w celu zapewnienia swojej działalności w trybie zdalnym. Wprowadzono rozwiązania elektroniczne, wymagające wykorzystanie odpowiedniego sprzętu (kamery internetowe, mikrofony, łącze internetowe o odpowiedniej przepustowości). Studenci i pracownicy nabyli nowe kompetencje. Wprowadzone rozwiązania zostaną najprawdopodobniej wykorzystywane również po ustaniu pandemii.

Wśród wymienianych zalet transformacji cyfrowej można wyróżnić (Hai et al. (2021)):

- Zastąpienie tradycyjnych przepływów pracy przez cyfrowe procesy
- Zwiększenie efektywności pracy i minimalizacja błędów technicznych
- Szybkie oraz elastyczne wdrażanie nowych usług i technologii
- Zwiększenie jakości pracy oraz wydajności

Cyfrowa transformacja nie jest jednopunktowym procesem, a raczej ruchomym celem, za którym firmy instytucje MŚP powinny podążać w celu zapewnienia swojej konkurencyjności na rynku (Priyono et al. 2020). Transformacja cyfrowa nie ogranicza się jedynie do swojej technologicznej części, ale potrzebuje udziału ludzi, bowiem jest dla ludzi (Fonseca et al.(2021)). COVID-19 zweryfikował potrzeby użytkowników oraz dostarczanych usług w cyberprzestrzeni. Najnowsze raporty „Data Never Sleeps” 8.0 („Domo Resource - Data Never Sleeps 8.0”, <https://www.domo.com/learn/infographic/data-never-sleeps-8> (2021)) i 9.0 („Domo Resource - Data Never Sleeps 9.0”, [www.domo.com/learn/infographic/data-never-sleeps-9](https://www.domo.com/learn/infographic/data-never-sleeps-9) (2022)) z

odpowiednio 2020 i 2021 roku obrazują statystyczne wykorzystanie danych lub wartość zakupów w cyberprzestrzeni w „każdej minucie” podając czołowe aplikacje oraz wolumen danych przetwarzanych w każdej minucie doby. W trakcie pandemii zauważono wyraźny wzrost usługodawców świadczących usługi zdalne, np. platformy do zamawiania jedzenia, pracy w zespołach, platformy do zdalnego nauczania lub rozrywki online. W trakcie pandemii spadło zainteresowanie platformami do usług przewozu osób, portali randkowych i rezerwacji noclegów online („Domo Resource - Data Never Sleeps 7.0”, <https://www.domo.com/learn/infographic/data-never-sleeps-7> , (2020)). Pandemia COVID-19 wymusiła rozwój społeczeństwa informacyjnego na świecie. Z ostatnio publikowanego raportu wynika wzrost populacji „internetowej” w 2021 roku do 5,2 miliarda osób, w porównaniu do 4,3 miliarda w 2018 roku oraz 4,5 miliarda w 2020 roku.

### 1.5.3 Stosowanie obostrzeń

Wiele krajów zdecydowało się na wprowadzenie restrykcji mających na celu spowolnienie rozprzestrzeniania się wirusa SARS-CoV-2 oraz przeciążanie systemu opieki zdrowotnej. Zdecydowano się nawet na wprowadzenie lockdown’ów i innych ekstremalnych restrykcji, których luzowanie budziło obawy zdrowotne, społeczne i ekonomiczne.

Cheval et al. (2020) wskazuje na Włochy jako jedno z pierwszych w Europie dotkniętych pandemią oraz modelowe państwo w do badań hipotez w pierwszych miesiącach pandemii. Badano wpływ klimatu, promieniowania UV i temperatury na rozprzestrzenianie się wirusa. Klimat Włoch jest zróżnicowany: na północy chłodny z najsuchszą od 60 lat zimą w 2020 roku, natomiast cieplejszy w południowej części. Poszczególne regiony Włoch wprowadziły zróżnicowane polityki ograniczeń mające na celu ograniczenie rozprzestrzeniania się wirusa. Trwogę budziły potencjalne drogi rozprzestrzeniania się wirusa oraz wpływ na społeczne i ekonomiczne aspekty państwa oraz na skalę kontynentu europejskiego.

Kędziński et al. (2020) zaproponował wielowymiarowy model analityczno-decyzyjny związany z polityką zarządzania restrykcjami. Wskazywał, restrykcje były wprowadzane pod presją czasu i wykładniczo rosnącej liczby zakażeń. Kierowano się jedynie kryteriami medycznymi jednocześnie pomijając konsekwencje społeczno-gospodarcze. W przypadku COVID-19 restrykcje koncentrowały się wokół:

- ograniczeń w przemieszczaniu się
- edukacji i instytucji kultury
- opieki nad dziećmi
- ograniczenia dużych zgromadzeń publicznych
- komunikacji publicznej
- działalności placówek handlowych i usługowych

Han et al. (2020) zbadał podejścia 9 wysokodochodowych krajów i regionów luzujących restrykcje COVID-19. Wśród analizowanych restrykcji były ograniczenia związane z przekraczaniem granic, śledzenie kontaktów (na podstawie danych z systemów informatycznych, aplikacji mobilnych, bluetooth, transakcji kartami płatniczymi, skanowania kodów QR, położenie GPS telefonów komórkowych), kwarantanna i dystans społeczny.

Izraelski rząd upoważnił służby specjalne do monitorowania zachowania swoich obywateli za pomocą cyfrowych mobilnych urządzeń (Tidy, J. (2021). Coronavirus: Israel enables emergency spy powers. 2020. URL <https://www.bbc.com/news/technology-51930681>). Działania na pograniczu praw obywatelskich wprowadzono w trybie nadzwyczajnym na ograniczony 30-dniowy okres czasu, ale umożliwiło to śledzenie przestrzegania kwarantanny, identyfikację osób mających wcześniejszy kontakt z zakażonym oraz zbudowanie modelu do predykcji. Model bazował na SVR i danych klinicznych ponad 40,5 tys. pacjentów uzyskując błąd na poziomie 0,87-3,51% dla prognozy jednodniowej, 1,02-5,63% dla prognozy 3-dniowej oraz 0,95-6,90% dla prognozy ponad sześciodniowej (Lalmuanawma et al. (2020)).

Alenzi et al. (2021) w swoich badaniach porównujących modele autoregresyjne, SIR i SEIR stwierdził, że wprowadzone środki zapobiegawcze jak ograniczenia w przemieszczaniu się oraz lockdown były uzasadnione w celu kontroli rozprzestrzeniania się COVID-19. Inne środki ostrożności (dystans społeczny i noszenie masek zakrywających usta i nos) były pomocne w ograniczeniu rozprzestrzeniania się choroby.

#### 1.5.4 Kwarantanna

Ludzkość w średniowieczu zaczęła wyciągać wnioski z poprzednich pandemii. Zaczęto stosować powszechnie środki zapobiegawcze, którym była np. kwarantanna. Średniowieczne społeczeństwa doświadczane licznymi pandemiemi zaobserwowały związek pomiędzy wpływem

czasu a zaobserwowaniem symptomów choroby (Huremović, 2019). Twierdzono, że skoro po okresie obserwacji danej osoby nie stwierdzono u niej objawów choroby, to prawdopodobieństwo zakażenia a tym samym rozprzestrzeniania choroby po wjeździe do danego miasta jest niskie. Pierwszą kwarantannę odnotowano w Ragusie (Chorwacja) w 1377 r. Wszyscy przyjezdni przed wjazdem do miasta byli zobowiązani do spędzenia 30 dni na okolicznej wyspie Lorkum. 30-dniowy okres (zw. *trentine*) został wydłużony do 40 dni (*quarantine*) (Sehdev, 2002).

Kwarantanna wiąże się z ograniczeniem poruszania się lub oddzieleniem od reszty populacji osób zdrowych. Oddzielane są osoby narażone na kontakt z wirusem w celu monitorowania objawów i zapewnienia wczesnego wykrywania przypadków zakażenia (WHO, (2020)). Współcześnie kwarantanna jest uregulowana prawnie w IHR (2020) artykułach 30, 31 i 32.

### 1.5.5 Izolacja

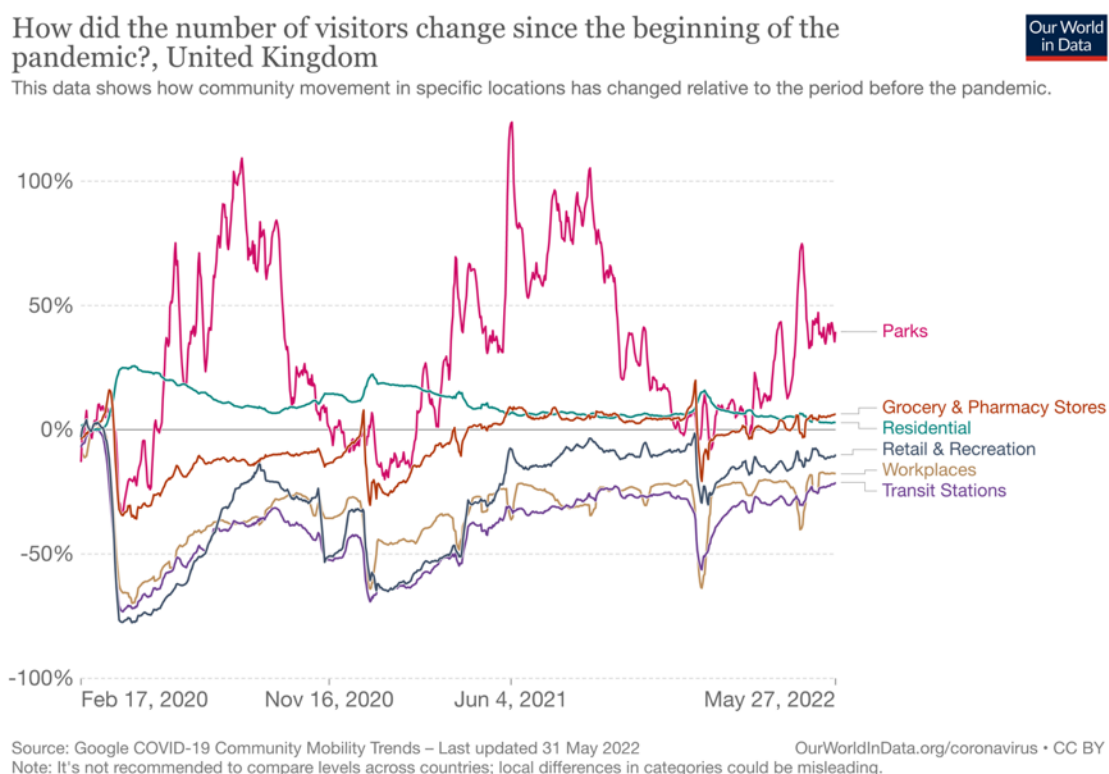
Jednym ze sposobów walki z pandemią była izolacja osób zakażonych. W czasach pandemii Czarnej śmierci chorych izolowano w zamkniętych pomieszczeniach zwanych lazaretami. Podobne metody stosowano w pierwszej pandemii cholery [Poiret i Boivin (2021)].

Izolacja różni się od kwarantanny tym, że izolacja to separacja osoby chorej lub zarażonej w celu zapobiegania rozprzestrzenianiu się infekcji (WHO, (2020)). Izolacja może się odbywać w domu chorego jak i w placówkach służby zdrowia.

### 1.5.6 Ograniczenie mobilności

Ograniczenie mobilności jest jednym z najpowszechniej stosowanych obostrzeń. W trakcie pierwszej pandemii cholery w Indiach wprowadzono obostrzenia polegające na odmowie wstępu do portów statków przybywających z regionów dotkniętych cholera [Poiret i Boivin (2021)]. Na potrzeby badań naukowych nad COVID-19 firma Google udostępniła raporty mobilności bazujące na danych pochodzących z urządzeń mobilnych. Mobilność jest mierzona na podstawie kategorii odwiedzanych miejsc: handel i rekreacja, sklepy spożywcze, apteki, miejsca pracy oraz stacje przesiadkowe. Dane są zbierane anonimowo za zgodą użytkowników – posiadaczy kont Google ("Google COVID-19 Community Mobility Reports". <https://www.google.com/covid19/mobility/> (2022)). Analiza przemieszczania się oraz związane z nim ograniczenia umożliwiają tworzenie modeli badających wpływ mobilności na

rozprzestrzenianie się wirusa, co może mieć kluczowe znaczenie dla decydentów (Ivari et al. (2020)). Wizualizacje danych pozyskanych z "Google COVID-19 Community Mobility Reports" pokazują jak w trakcie trwania pandemii zmieniały się zachowania społeczne - częstotliwość odwiedzania miejsc publicznych (Rys. 2. ), wychodzenia z domu w porównaniu z czasem sprzed pandemii (Rys. 3. ) oraz ograniczenia w przekraczaniu granic (Rys. 4. ).

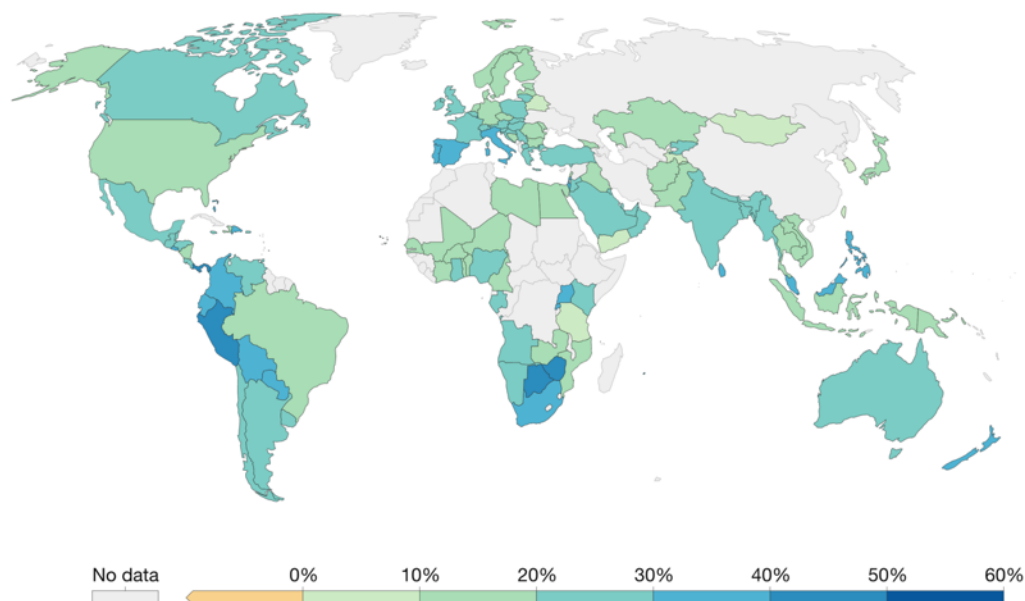


Rys. 2. Zmiana odwiedzin poszczególnych kategorii miejsc w Wielkiej Brytanii od początku pandemii. Graficzna interpretacja danych z Google COVID-19 Community Mobility Reports („COVID-19: Google Mobility Trends”, <https://ourworldindata.org/covid-google-mobility-trends> (2022))

## Residential areas: How did the time spent at home change relative to before the pandemic?, Apr 15, 2020



This data shows how the number of visitors to residential areas has changed relative to the period before the pandemic.



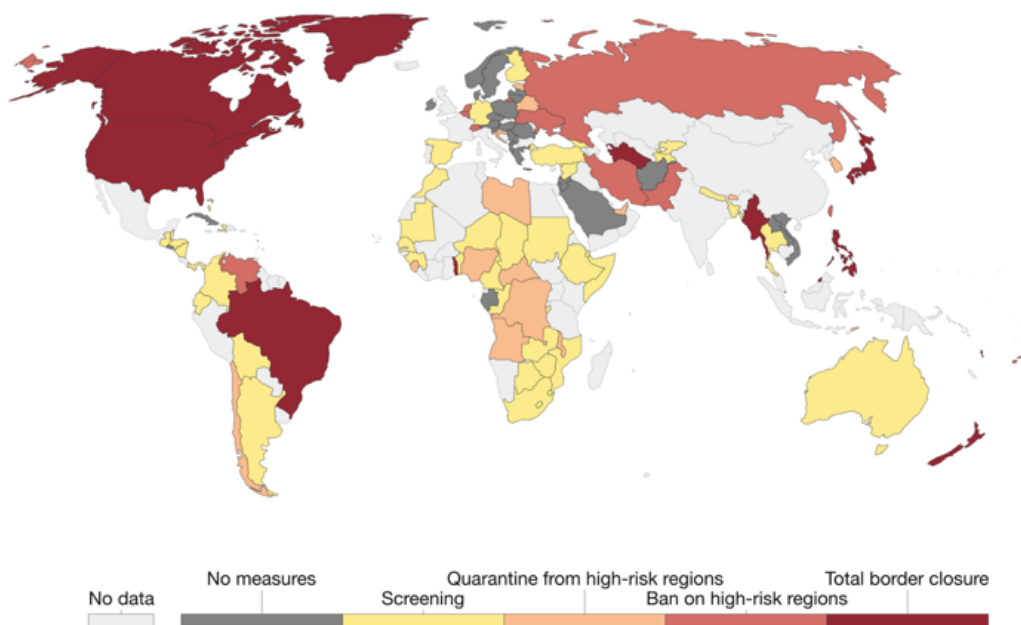
Source: Google COVID-19 Community Mobility Trends – Last updated 31 May 2022

OurWorldInData.org/coronavirus • CC BY

Note: It's not recommended to compare levels across countries; local differences in categories could be misleading.

*Rys. 3. Zmiana czasu spędzanego (świat z podziałem na kraje) w domu od wybuchu pandemii do 15 kwietnia 2020 roku w stosunku do czasu sprzed pandemii. Graficzna interpretacja danych z Google COVID-19 Community Mobility Reports („COVID-19: Google Mobility Trends”, <https://ourworldindata.org/covid-google-mobility-trends> (2022)).*

Najbardziej dotknięty ograniczeniami mobilności był ruch lotniczy. Ograniczenia związane z podróżowaniem spowodowały spadek ruchu lotniczego o 41,7% ogółu lotów w 2020 r. („Commercial flights down 42% in 2020”, <https://www.flightradar24.com/blog/commercial-flights-down-42-in-2020/> (2021)).



Source: Oxford COVID-19 Government Response Tracker, Blavatnik School of Government, University of Oxford – Last updated 1 June 2022  
OurWorldInData.org/coronavirus • CC BY

Rys. 4. Zobrazowanie zróżnicowania światowego podejścia do kontroli sanitarnej dla osób wjeżdżających do danego kraju. Graficzna interpretacja danych z Google COVID-19 Community Mobility Reports („COVID-19: Google Mobility Trends”, <https://ourworldindata.org/covid-google-mobility-trends> (2022)).

### 1.5.7 Komunikacja społeczna i jakość danych w dobie kryzysu

Harris i Moss (2020) wskazują na wyzwanie wynikające z pandemii COVID-19 dotyczące trudności w kontrolowaniu populacji bez wsparcia systemów jakości oraz danych. Skutkuje to wprowadzaniem krótkotrwałych polityk. W dłuższej perspektywie pandemia wpływa na obniżone morale społeczeństwa oraz przekłada się na brak zaufania do rządu. Dlatego spójne zarządzanie kryzysem pandemicznym zarówno w perspektywie krótko- jak i długoterminowej oraz komunikacja są krytyczne.

Istotną rolę w komunikacji społecznej odegrały media społecznościowe. Rządy i instytucje za pomocą tych kanałów informacyjnych przekazywały dane dotyczące liczb związanych z nowymi przypadkami zakażeń, zgonami, szczepieniami, testami. Media społecznościowe również były wykorzystywane do komunikowania zaleceń, obostrzeń oraz jako platforma do wymiany

opinii, pomysłów, lęków, teorii spiskowych itp. związanych z walką z COVID-19 (Tsoy et al. (2021)).

Moein et al. (2021) wskazywał na niedostatek danych w początkowej fazie pandemii do opracowywania złożonych modeli, bowiem dotychczas stosowane powszechnie (SIR) nie odpowiadały specyfice COVID-19. Wskazywał, że precyzyjne dane epidemiologiczne i biomedyczne są niezbędne do przewidywania rozwoju pandemii.

#### 1.5.8 Spójność danych statystycznych na przykładzie Polski

Walkowiak i Walkowiak (2022) wskazują, że w czasie drugiej i trzeciej fali COVID-19 liczba zgonów w Polsce w oficjalnych statystykach była niedoszacowana, stanowiąc tylko 60 % rzeczywistych przypadków śmiertelnych na podstawie zaproponowanego modelu. W przeciwieństwie do oficjalnych komunikatów zrównujących śmiertelność drugiej i trzeciej fali, stwierdzono w wyniku badań nadmierną śmiertelność drugiej fali przewyższającą znacznie śmiertelność trzeciej fali. Po zestawieniu czasu pomiędzy nadmierną liczbą zgonów w różnych grupach demograficznych oraz zgonami oficjalnie zarejestrowanymi ze wskazaniem przyczyny jako COVID-19 a ogólną liczbą zgonów stwierdzono, że co najmniej 95% zgonów z ogółu musiało być spowodowanych bezpośrednio przez COVID-19.

Z początkiem kwietnia 2022 roku Ministerstwo Zdrowia ogłosiło zmianę raportowania danych o COVID-19. Zaprzesano podawania informacji o kwarantannie oraz testowaniu. Przyczyną było ogólne zniesienie obostrzenia w postaci kwarantanny. Zlikwidowano również tzw. „covidowe” szpitale i oddziały szpitalne, a co za tym idzie zaprzestano też raportowania informacji o obłożeniu łóżek w szpitalach związanych z COVID-19. Zmieniono zasady testowania, utrzymano testowanie antygenowe w POZ. Uznano, że będzie stała liczba zakażeń utrzymująca się co roku na podobnym poziomie, czyli następuje przejście z epidemii do endemii. Chorobę COVID-19 porównano pod tym względem do epidemii grypy.

## 1.6 Krajowe zespoły eksperckie

W obliczu wyzwań związanych z pandemią COVID-19, nieocenioną rolę odegrały zróżnicowane środowiska naukowe, włączając w to zespoły eksperckie zainicjowane przez władze krajowe oraz samodzielnie działających naukowców. Wśród nich znaleźli się przedstawiciele

takich grup jak Interdyscyplinarne Centrum Modelowania Uniwersytetu Warszawskiego, zespół doradczy ds. monitorowania i prognozowania epidemii przy Ministrze Zdrowia, Departament Analiz i Strategii Ministerstwa Zdrowia, a także Grupa MOCOS z Politechniki Wrocławskiej. Każda z tych jednostek wniosła własny, niepowtarzalny wkład w analizę i prognozowanie przebiegu pandemii, pracując nad modelami epidemiologicznymi i oceną wpływu wprowadzanych ograniczeń. Ich działania były kluczowe dla zrozumienia mechanizmów rozprzestrzeniania się COVID-19 i kształtowania strategii reagowania na pandemię, dostarczając niezbędnych informacji zarówno decydującym, jak i publicznym instytucjom zdrowia.

### 1.6.1 Zespół modelu epidemiologicznego Interdyscyplinarnego Centrum Modelowania Uniwersytetu Warszawskiego

Zespół modelu epidemiologicznego Interdyscyplinarnego Centrum Modelowania Uniwersytetu Warszawskiego wchodzi w skład grupy doradczej Ministerstwa Zdrowia i udostępnia wyniki predykcji w krótkiej skali czasowej oraz warianty rozprzestrzeniania się epidemii w dłuższych perspektywach z uwzględnieniem odpowiednich obostrzeń lub parametrów choroby.

Zespół opracował Model Epidemiologiczny ICM, który opisuje rozwój epidemii COVID-19 w Polsce. Model ten ma za zadanie predykcję potencjalnych ścieżek rozwoju epidemii oraz badanie wariantów scenariuszy i ich efektów stosowanych restrykcji administracyjnych. Model Epidemiologiczny ICM wywodzi się z lat 2008-2010 i był wykorzystywany do symulacji rozprzestrzeniania się wirusa grypy. Model umożliwia śledzenie rozwoju epidemii na siatce geograficznej o rozdzielczości 1 km<sup>2</sup> oraz kontakty fizyczne uznane za drogę przenoszenia wirusa z rozróżnieniem na kategoryzację miejsc transmisji (np. gospodarstwa domowe, miejsca pracy, miejsca edukacji oraz przypadkowe spotkania uliczne).

Zespół podkreśla na swojej witrynie internetowej, że musi się zmagać ze zdobywaniem danych niezbędnych do konstrukcji i kalibracji modelu. Kluczowymi parametrami są bazowa transmisyjność wirusa oraz skłonność do wychodzenia z domu przy lekkich objawach, kalibrowane po bayesowsku do rzeczywistego przebiegu pandemii [„Covid 19 - ICM UW” <https://covid-19.icm.edu.pl/> (2022)].

### 1.6.2 Zespół ds. monitorowania i prognozowania epidemii COVID-19 powołany przez Ministra Zdrowia

Minister Zdrowia powołał zespół ds. monitorowania i prognoz, w skład którego wchodzi zespoły ekspercko badawcze oraz przedstawiciele instytucji publicznych:

- Ministerstwa Zdrowia,
- Głównego Inspektoratu Sanitarnego,
- Narodowego Instytutu Zdrowia Publicznego – Państwowego Zakładu Higieny,
- Głównego Instytutu Statystycznego,
- Agencji Oceny Technologii Medycznych i Taryfikacji,
- Centrum e-Zdrowia

Celem powołania takiego zespołu jest wymiana doświadczeń, porównywanie danych różnych klas modeli epidemiologicznych i podejść do prognozowania epidemii. Informacje o zespole umieszczone są na stronie internetowej Ministerstwa Zdrowia [„Zespół Ministra Zdrowia ds. monitorowania i prognoz - Koronawirus: informacje i zalecenia - Portal Gov.pl” <https://www.gov.pl/web/koronawirus/zespol-ministra-zdrowia-do-spraw-monitorowania-i-prognoz> (2022)]. Zespół zajmuje się zbieraniem zapotrzebowania i udostępniania danych zespołom badawczym, opracowywaniem krótko- i średnioterminowych prognoz rozwoju epidemii i prognoz wariantowych. Swoje analizy opiera na danych związanych z liczbą zakażeń, zgonów, testów itp. Dane te są podstawą rządu do bieżącego reagowania, wprowadzania zasad i obostrzeń odzwierciedlających sytuację epidemiologiczną oraz różne scenariusze rozwoju. Brak jest informacji odnośnie daty powołania takiego zespołu, wykorzystywanych modeli, porównania modeli oraz wizualizacji danych. Autor nie znalazł danych kontaktowych do zespołu. Nie umieszczono ich na stronie internetowej.

### 1.6.3 Departament Analiz i Strategii Ministerstwa Zdrowia

Ogólnodostępna strona internetowa Departamentu Analiz i Strategii Ministerstwa Zdrowia nie zawiera informacji o zadaniach związanych stricte z pandemią COVID-19, natomiast są to zadania związane z powszechną służbą zdrowia. [„Departament Analiz i Strategii - Ministerstwo Zdrowia - Portal Gov.pl” <https://www.gov.pl/web/zdrowie/departament-analizstrategii> (2022)]

#### 1.6.4 Grupa MOCOS z Politechniki Wrocławskiej

Grupa MOCOS (MOdelling COronavirus Spread) to międzynarodowy interdyscyplinarny zespół naukowców zajmujący się modelowaniem epidemii COVID-19.

Informacje o działalności grupy są dostępne na stronie jej internetowej [„MOCOS : Strona Główna” <https://mocos.pl/pl/> (2022)]. Grupa publikuje krótkoterminowe prognozy rozwoju pandemii, analizy i symulacje, aplikację do monitorowania wariantów SARS-CoV-2 w Polsce oraz raporty dotyczące stanu epidemii.

Model stosowany przez grupę MOCOS jest częścią „German & Polish COVID-19 Forecast HUB”, który udostępnia prognozy rozwoju epidemii na najbliższe 4 tygodnie. Model mikrosymulacyjny symuluje zachowania indywidualnych agentów. Model ten umożliwia dokonanie szeregu analiz (Krueger (2020)).:

- aktualnego poziomu reprodukcji wirusa
- etapu rozwoju epidemii
- skali ograniczenia kontaktów międzyludzkich
- poziomu wykrywalności zarażonych wirusem
- wpływu rekomendowanych działań na możliwości rozwoju epidemii

Członkowie grupy opracowali aplikację internetową estymującą indywidualne ryzyko hospitalizacji lub zgonu bazując na interaktywnym formularzu ankiety [“COVID-19 Individual Risk Assessment” <https://crs19.pl/> (2022)]. Model bazuje na przeszło 52 tys. przypadków zakażeń (pozytywny wynik testu PCR) z frakcji populacji Polski. Algorytm modelu bazuje na tree boosting „wzmocnienie drzewa” (biblioteka xgboost). Celem symulacji jest uświadomienie użytkownika o rosnącym wraz z wiekiem i chorobami współistniejącymi ryzyku hospitalizacji i zgonu.

### 1.7 Powszechnie stosowane metody predykcji rozprzestrzeniania się wirusów

Pojawienie się pandemii SARS-CoV-2 wywołało ewidentny wybuch zainteresowania epidemiologicznych i matematycznych modeli postępowania choroby. Naukowcy podejmowali próby zastosowania klasycznych modeli epidemiologicznych, takich jak niemal stuletni SIR lub model oparty na agentach, modeli ekonometrycznych (AR i ARIMA), statystycznych (regresja)

oraz stosować uczenie maszynowe. Przyjęcie założeń bazujących na doświadczeniach z wcześniej poznanymi wirusami lub chorobami jest oderwane od specyfiki COVID-19. Przykładowo zestawiając model dla wścieklizny z COVID-19 zakłada się nieuniknioną śmierć po zakażeniu, pomija się możliwość ozdrowienia i powtórnego zakażenia, nie wspominając już o mutacjach wirusa. Badacze na początku pandemii wychodzili w swoich analizach z klasycznych modeli zwracając uwagę na ich niedoskonałości (Moein et al. (2021)), jednocześnie rozbudowując modele klasyczne lub tworząc modele hybrydowe. Wszystkie modele oparte są na dostępnych danych. Dane związane z COVID-19 są zbiorem szeregów czasowych. Szeregiem czasowym nazywamy dane zbierane w określonych interwałach czasowych. Analiza szeregów czasowych na podstawie danych historycznych umożliwia znalezienie wzorców, trendów, podobieństw, cech sezonowych itp. z przeszłości oraz prognozowanie przyszłych wartości. Dane COVID-19 mają charakter dynamiczny, więc stosowanie modeli statystycznych i epidemiologicznych z góry przyjętymi założeniami i liniowością dają nieprecyzyjne wyniki (Chimmula et al. (2020)).

Poniżej przedstawiono metody, z którymi zetknięto się w czasie przygotowywania niniejszej rozprawy doktorskiej. Metody te ze względu na swoją popularność wśród badaczy nie stanowią głównego przedmiotu badań, niemniej jednak są rozpatrywane jako jedne ze sposobów predykcji. Przegląd literatury zawiera opis metod wspierany równaniami. Wskazano wnioski z badań związanych z prognozą w szeregach czasowych pandemii z wykorzystaniem danych metod. Wskazano stosowane metryki do porównywania dokładności prognoz.

Przegląd otwiera klasyczny model SIR oraz jego pochodne. Przedstawiono podstawowe równania związane z tym modelem. Zestawiono również wyniki predykcji wybranymi metodami oraz wskazano bibliotekę CovsirPhy jako popularne narzędzie do modelowania pandemii z wykorzystaniem modelu SIR i jego pochodnych. Kolejnym analizowanym modelem jest model matematyczny oparty na agentach oraz jego adaptacja dokonana przez Bhatnagar'a. Wskazano jego ograniczenia w kontekście COVID-19. W przeglądzie literatury nie mogło zabraknąć powszechnie stosowanego modelu ARIMA oraz jego hybrydowego zastosowania z pochodnymi SIR oraz porównanie z prognozowaniem za pomocą ANN. Przegląd literatury metod modelowania związanego z pandemią zamyka przegląd metod DL oraz ich porównanie.

### 1.7.1 SIR i metody pochodne

Model SIR został zaproponowany w 1927 roku w celu symulacji rozprzestrzeniania się chorób zakaźnych takich jak odra i różyczka [Kermack & McKendrick (1927)]. W późniejszych latach SIR był stosowany w celu analizy i symulacji nowych wirusów, takich jak Ebola (Bhandari (2019)), HIV (Alenezi et al. (2021)), MERS [Chang (2017), Eifan et al. (2017)]. Model SIR stosowany jest również w weterynarii (de Jong (1995)), na przykład Gerasi et al. (2019) wykorzystał SIR do badania rozprzestrzeniania się choroby zakaźnej ASF. Model ten w swojej pierwotnej formie wymagał niewielkiej ilości wiarygodnych danych epidemiologicznych i uznawany jest za model klasyczny w epidemiologii.

Model SIR był adoptowany na potrzeby badania wirusów do nowych modeli. COVID-19 analizowano pochodnymi modelami od SIR z uwzględnieniem dodatkowych czynników właściwych dla tej choroby. Takie modele stworzyli de León et al. (2020), Ndaïrou et al. (2020) oraz Samui et al. (2020).

Kumari et al. (2021) wskazał niedoskonałości modelu SEIJR wykorzystywanego do badania wybuchu epidemii SARS, który okazał się niewłaściwym modelem w stosunku do COVID-19. Brakowało w tym modelu predykcji wpływu pandemii ze względu na brak istotnych dla COVID-19 współczynników jak kwarantanna oraz przypadki zakażeń bezobjawowych. Kumari et al. (2021) zaproponował bardziej rozbudowany model SEIAQRDT, wykorzystany do badania rozprzestrzeniania się COVID-19 na przykładzie Indii. Nazwa modelu pochodzi od angielskojęzycznych akronimów określających enumeratywnie dopuszczalne stany osobników z populacji: podatne, narażone, zakażone, bezobjawowe, poddane kwarantannie, wyleczone, martwe i niewrażliwe. Model SEIAQRDT uwzględniał podział na przypadki bezobjawowe i objawowe, ogólnokrajowy lockdown oraz wpływ noszenia masek zakrywających nos i usta.

De León et al. (2022) zaproponował model SEIARD uwzględniający szczepienia na przykładzie USA. Model matematyczny zawiera dwa oddzielne zestawy równań dla szczepów wirusa SARS-CoV-2, które wpływają na tę samą populację i jej zaszczepioną subpopulację.

Model SEIAR różni się od SEIJR brakiem izolowanych osobników, w zamian uwzględnia osobniki bezobjawowe (Huang et al. (2020)).

Model SIDARTHE jest jednym z najbardziej rozbudowanych modeli pochodnych od SIR. Wyróżnia spośród frakcji zarażonych kolejne frakcje: I - zakażonych, D-zdiagnozowanych, A-

chorych (zakażonych z nieokreślonymi objawami), R-rozpoznanych (zainfekowanych z symptomami), T-zagrożonych (zainfekowanych, z objawami zagrażającymi życiu). Badania wykonane za pomocą tego modelu potwierdzają istotę zmniejszenia współczynnika zakażeń w walce z COVID-19 poprzez stosowanie obostrzeń (Higazy (2020)).

Model SIR w swojej klasycznej formie prezentuje się następująco: liczba  $S$  oznaczająca podatne osobniki,  $I$  zakażające i  $R$  dla ozdrowiałych (i zarazem zmarłych - należących do tej samej frakcji w modelu), którzy swobodnie kontaktują się ze sobą ze współczynnikami przenoszenia choroby (zakażeń)  $\beta$  i ozdowień (i zarazem zgonów)  $\gamma$  w odosobnionej homogenicznej społeczności liczącej  $N$  - osobników, są ze sobą powiązane za pomocą następującego zestawu równań (Kudryashov et al. (2021)):

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \quad (1)$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

Założeniem modelu SIR jest niezmienna liczba osobników w populacji w czasie analizy i predykcji:

$$N = S + I + R \quad (4)$$

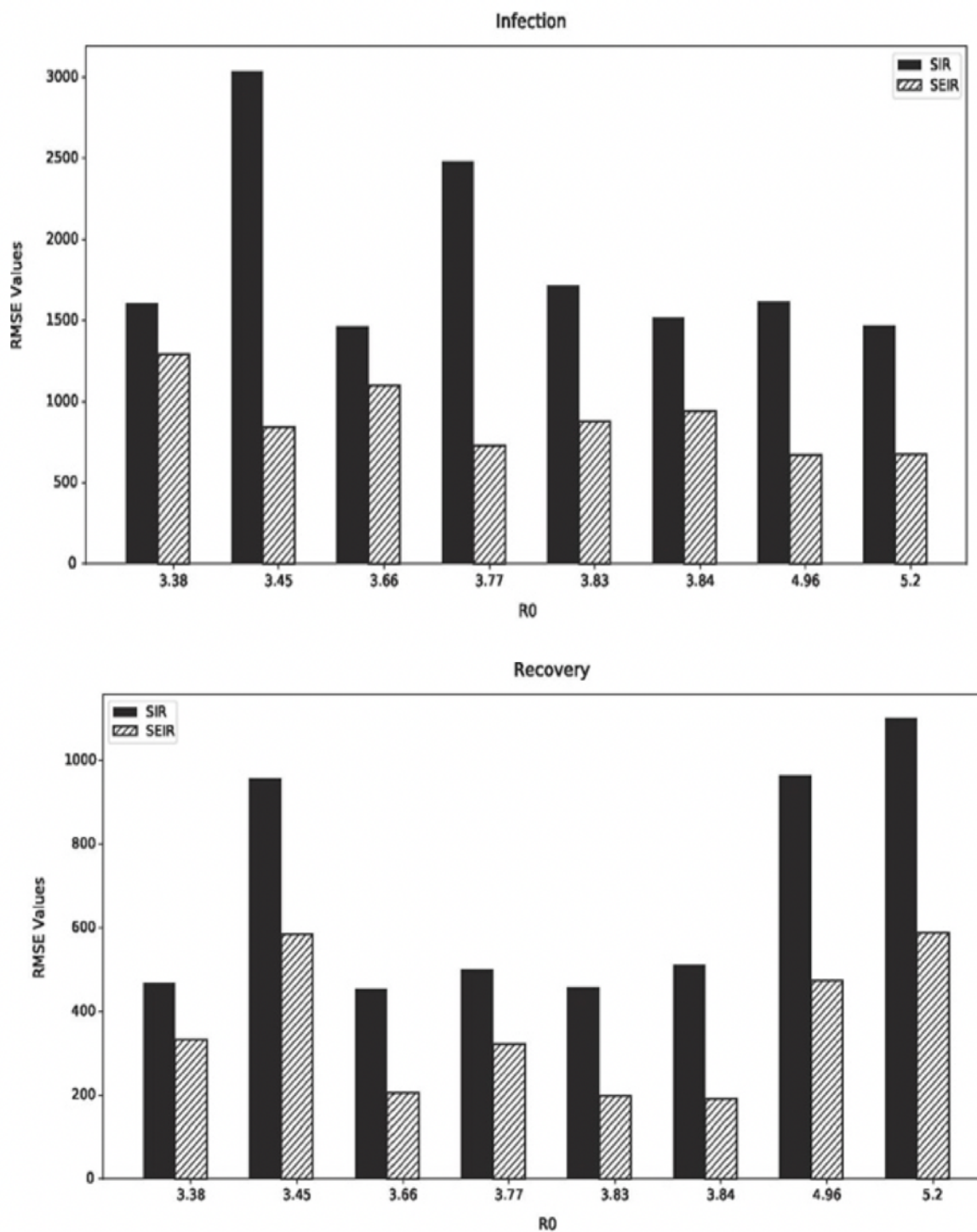
Oznacza to, że nie uwzględnia się nowonarodzonych osobników ani też zmarłych. Kluczowym parametrem w modelu jest liczba reprodukcji  $R_0$  (Alenezi et al. (2021)).

$$R_0 = \frac{\beta}{\gamma} \quad (5)$$

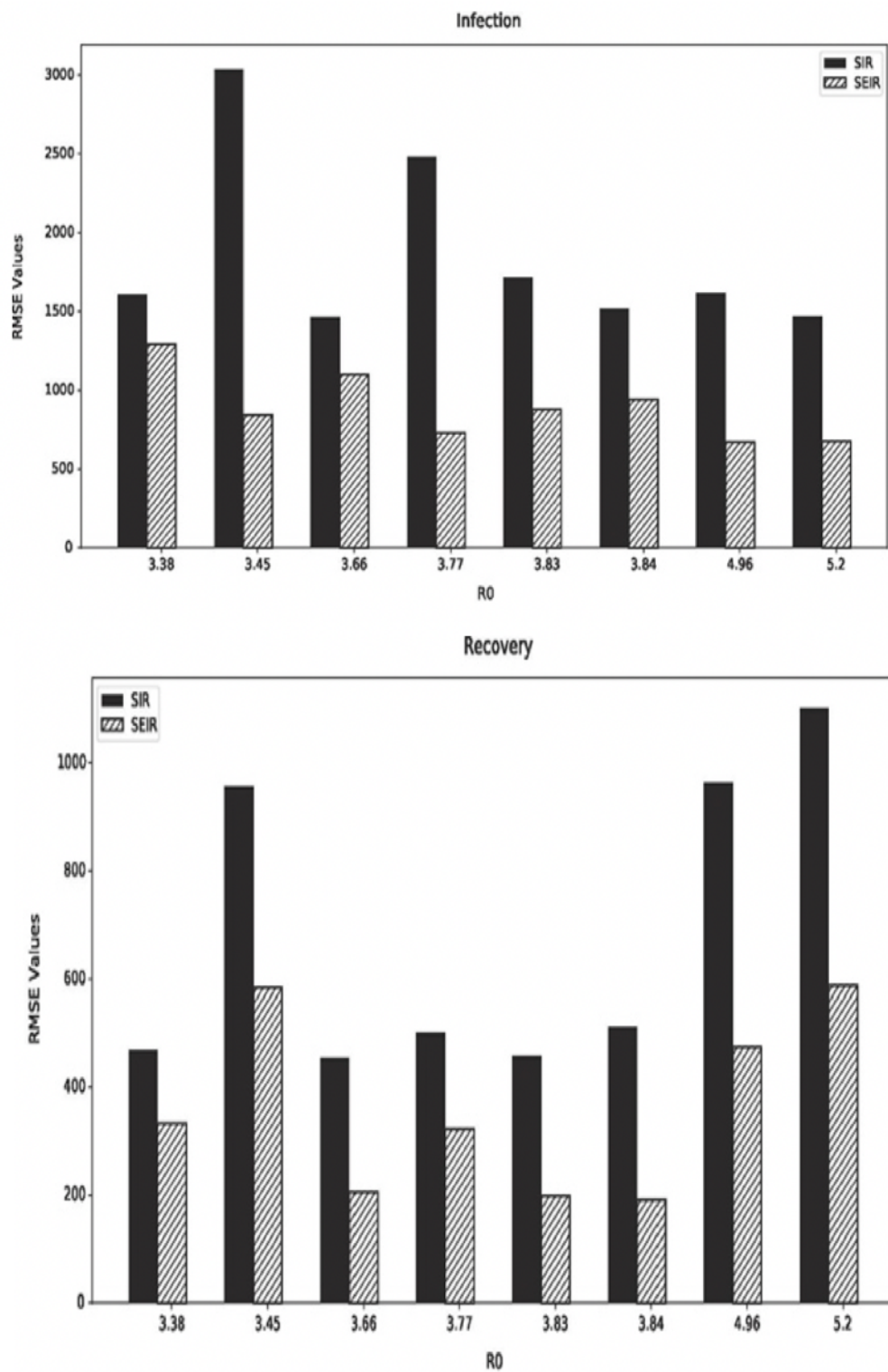
Copper et al. (2020) badał za pomocą modelu SIR różne istotne parametry rozprzestrzeniania się choroby COVID-19 w wybranych populacjach: Chin, Korei Południowej, Indii, Australii, USA, Włoch i stanu Texasu w USA. Moein et al. (2021) badał efektywność modelu SIR w predykcji COVID-19 populacji Isfahan w Iranie. Model SIR był stosowany do krótkoterminowej predykcji współczynników śmiertelności i wyzdowień w Wuhan w początkowej fazie pandemii. Stosowano ten model początkowo do predykcji skutków wprowadzanych polityk rządowych w Chinach oraz predykcji różnych strategii lockdown'u we Włoszech. Model sprawdzał się w krótkim horyzoncie predykcji, natomiast w długoterminowej predykcji nie potrafił przewidzieć rozprzestrzeniania się wirusa ani dostosować do wzorca choroby. Jego założenia okazały się

nieadekwatne do COVID-19 i była potrzeba rozwinięcia modelu z uwzględnieniem wiedzy biomedycznej i epidemiologicznych aspektów choroby.

Alenezi et al. (2021) porównał modele oparte na SIR i SEIR służące do predykcji rozwoju COVID-19. Porównanie predykcji zakażeń i ozdowień dla modeli SIR i SEIR przedstawiono na Rys. 5. i Rys. 6. Estymowane wartości  $R_0$  uzyskane za pomocą modelu SEIR są bliskie wartościom rzeczywistym, szczególnie w przypadku osób ozdowiających. Model SEIR okazał się znacznie precyzyjny w przypadku zakażeń i ozdowień.



Rys. 5. Porównanie wyników badań predykcji zakażeń za pomocą modeli SIR (czarny słupek) i SEIR (szary słupek) na podstawie RMSE. Źródło: Alenezi et al. (2021).



Rys. 6. Porównanie wyników badań predykcji ozdowień za pomocą modeli SIR (czarny słupek) i SEIR (szary słupek) na podstawie RMSE. Źródło: Alenezi et al. (2021).

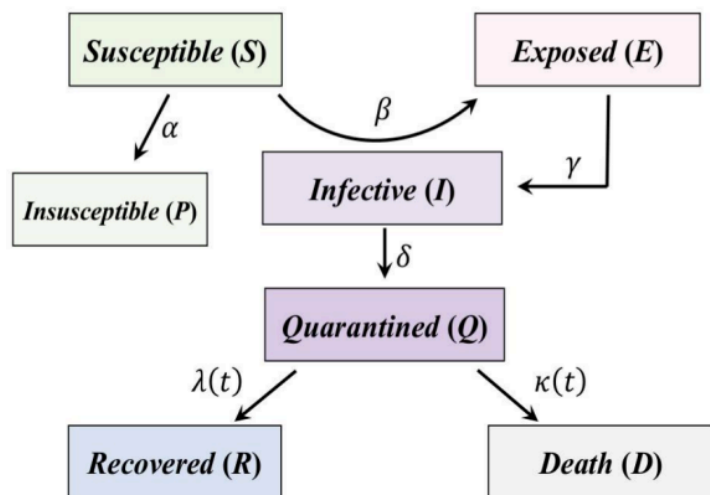
Peng et al. (2020) wskazuje, że model SEIR jest jednym z najszerzej rozbudowywanych modeli służących do charakterystyki wybuchu epidemii w Chinach oraz w innych krajach. Zmodyfikował klasyczny model SEIR wprowadzając 7 dodatkowych stanów (Rys. 7. ):

- S(t) – podatne przypadki
- P(t) – niepodatne przypadki
- E(t) – narażone przypadki (zakażone, ale jeszcze nie zakaźne)
- I(t) – zakaźne przypadki (posiadające zdolność zakaźną i jeszcze nie poddane kwarantannie)
- Q(t) – przypadki poddane kwarantannie (potwierdzone i zakażone)
- R(t) – ozdrowiałe przypadki
- D(t) – zamknięte przypadki (zgony)

Zależności między poszczególnymi stanami oraz charakterystyki wyrażone równaniami różniczkowymi przedstawiono na Rys. 7. Zmienna N wyrażona równaniem

$$N = S + P + E + I + Q + R + D \quad (6)$$

oznacza populację ograniczoną w danym zamkniętym regionie.<sup>44</sup>



### Generalized SEIR model

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta \frac{S(t)I(t)}{N} - \alpha S(t), \\ \frac{dE(t)}{dt} &= \beta \frac{S(t)I(t)}{N} - \gamma E(t), \\ \frac{dI(t)}{dt} &= \gamma E(t) - \delta I(t), \\ \frac{dQ(t)}{dt} &= \delta I(t) - \lambda(t)Q(t) - \kappa(t)Q(t), \\ \frac{dR(t)}{dt} &= \lambda(t)Q(t), \\ \frac{dD(t)}{dt} &= \kappa(t)Q(t), \\ \frac{dP(t)}{dt} &= \alpha S(t). \end{aligned}$$

Rys. 7. Model epidemiologiczny SEIR (zaznaczona część) z modyfikacją zaproponowaną przez Peng et al. (2020). Źródło: Peng et al. (2020).

Przedstawione współczynniki oznaczają odpowiednio:

- $\alpha$  - współczynnik ochrony
- $\beta$  - współczynnik infekcji
- $\gamma^{-1}$  - średni czas utajenia
- $\delta^{-1}$  – średni czas kwarantanny
- $\lambda(t)$  – współczynnik ozdowień
- $\kappa(t)$  – śmiertelność

Opracowanie kolejnych metod będących rozwinięciem metod pochodnych od SIR następuje wraz z poznaniem wirusa, określeniem poszczególnych stanów, średniego czasu przejść między stanami, np. od zakażenia do śmierci (Wiliński et al. (2022)).

W analizie COVID-19 za pomocą modelu SIR i pochodnych istotne jest wykorzystanie odpowiednich narzędzi. Szczególne miejsce zajmuje tutaj biblioteka CovsirPhy, której utworzenie zostało zainspirowane chęcią monitorowania rozprzestrzeniania się COVID-19, śledzenie danych na szczeblu regionalnym, krajowym i globalnym oraz badanie wpływu podejmowanych środków zapobiegawczych na rozprzestrzenianie się wirusa („covsirphy · PyPI”, <https://pypi.org/project/covsirphy/> (2022)).

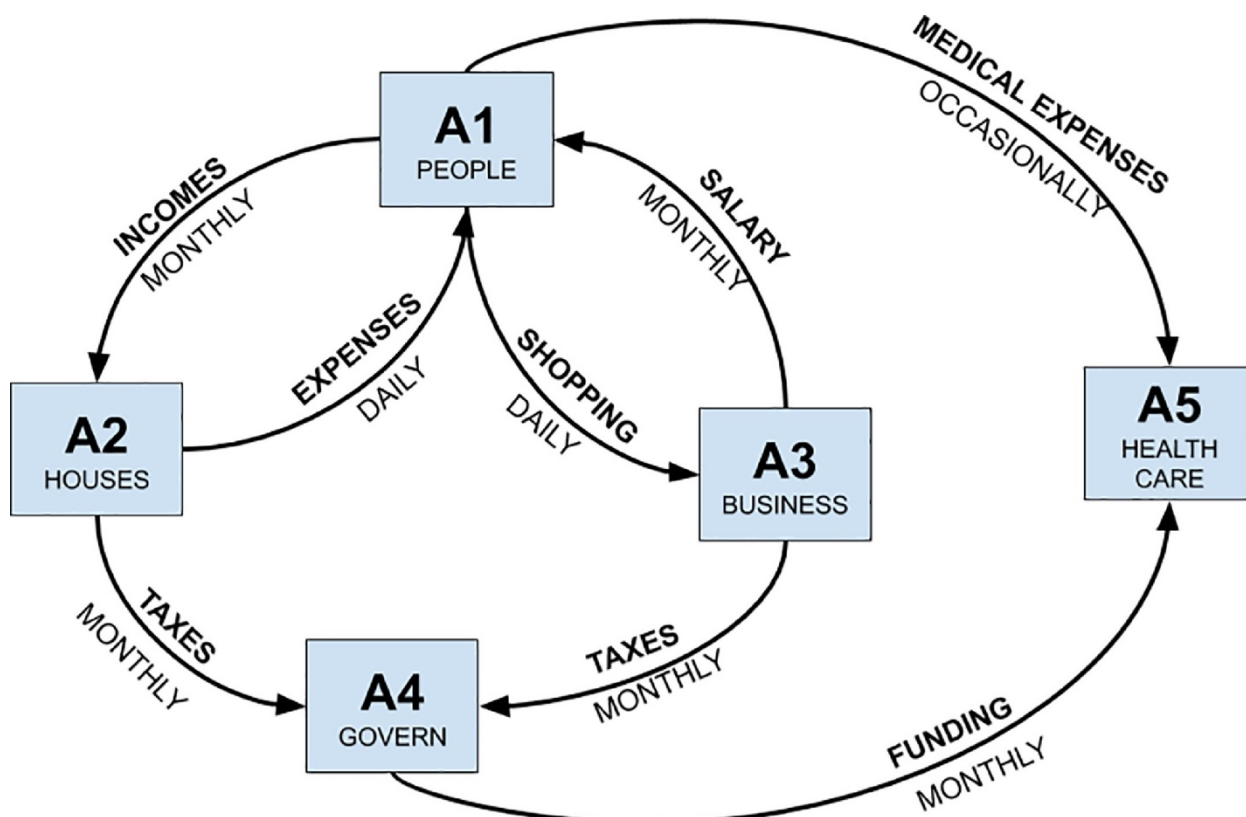
CovsirPhy jest biblioteką języka Python przeznaczoną do analizy danych związanych z COVID-19. Biblioteka ta wykorzystuje modele pochodne od SIR oparte na zwykłym równaniu różniczkowym. Dane są pobierane automatycznie z dostępnych online repozytoriów. Biblioteka posiada funkcjonalności do wizualizacji danych, symulacji modeli SIR, SIRD oraz SIRF (Werner et al. (2022)). Za pomocą wbudowanych funkcji istnieje możliwość zasymulowania dowolnych scenariuszy rozwoju pandemii, np. poprzez zasymulowanie wprowadzenia leku i szczepionki lub wpływu klimatu (Werner et al. (2021)). CovsirPhy umożliwia wykrycie zmian trendów w rozprzestrzenianiu się choroby, co jest przydatne do badania wpływu wprowadzanych obostrzeń lub zachowań społecznych na rozwój pandemii. Kolejną użyteczną funkcjonalnością jest możliwość predykcji przypadków choroby.

### 1.7.2 Model matematyczny oparty na agentach

Modele matematyczne oparte na agentach (ABM) są powszechnie wykorzystywanym narzędziem do predykcji zapotrzebowania na szpitalne łóżka w czasie pandemii COVID-19 (Hadley et al. (2022)).

Novakovic i Marshall (2022) wprowadził podejście do wykrywania punktów zmian w ABM (CP-ABM) w celu identyfikacji kluczowych zdarzeń prowadzących do zmian w kontaktach między ludźmi. Wykrywanie zmian w zachowaniu agentów umożliwiło osiągnięcie realistycznych wyników dla jednej populacji. Podejście pozwoliło na ocenę skuteczności wdrożonych nie-farmakologicznych środków zapobiegawczych takich jak regionalny lub krajowy lockdown oraz stosowanie maseczek ochronnych na usta i nos.

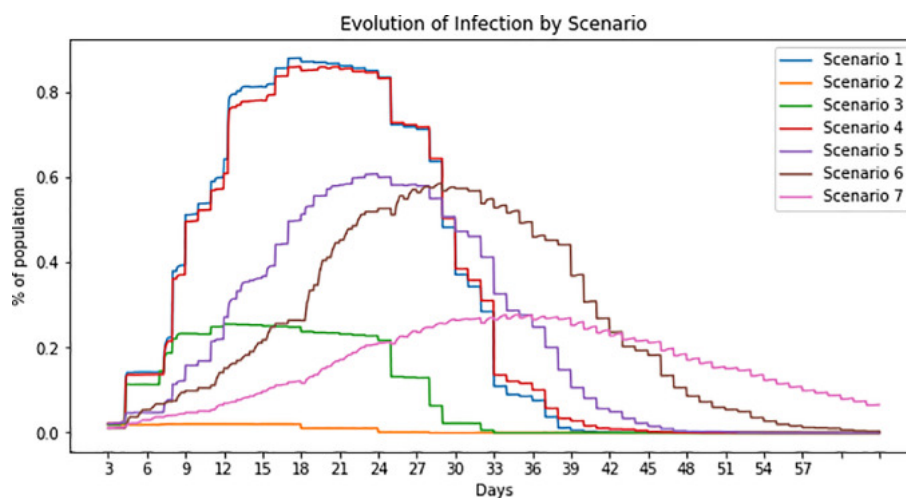
Silva et al. (2020) zaproponował nowy model SEIR symulujący dynamikę pandemii z wykorzystaniem społeczeństwa agentów emulujących ludzi, biznes, rząd i system służby zdrowia (Rys. 8. ).



Rys. 8. Podział na typ agentów oraz zależności ekonomiczne pomiędzy agentami w modelu proponowanym przez Silva et al. (2020). Źródło: Silva et al. (2020).

Rozważył 7 różnych scenariuszy dystansu społecznego z różnymi efektami epidemiologicznymi i gospodarczymi:

1. Brak aktywności
2. Lockdown
3. Warunkowy lockdown
4. Pionowa izolacja – zdiagnozowani zakażeni i osoby z grup ryzyka (osoby w podeszłym wieku z chorobami współistniejącymi) są w izolacji, natomiast ludzie młodzież i dorośli są nadal aktywni.
5. Częściowa izolacja
6. Użycie maseczek zakrywających usta i nos
7. Użycie maseczek zakrywających usta i nos w połączeniu z 50% izolacją społeczną.



Rys. 9. Analiza zakażeń dla poszczególnych scenariuszy modelu Silva et al. (2020). Źródło: Silva et al. (2020).

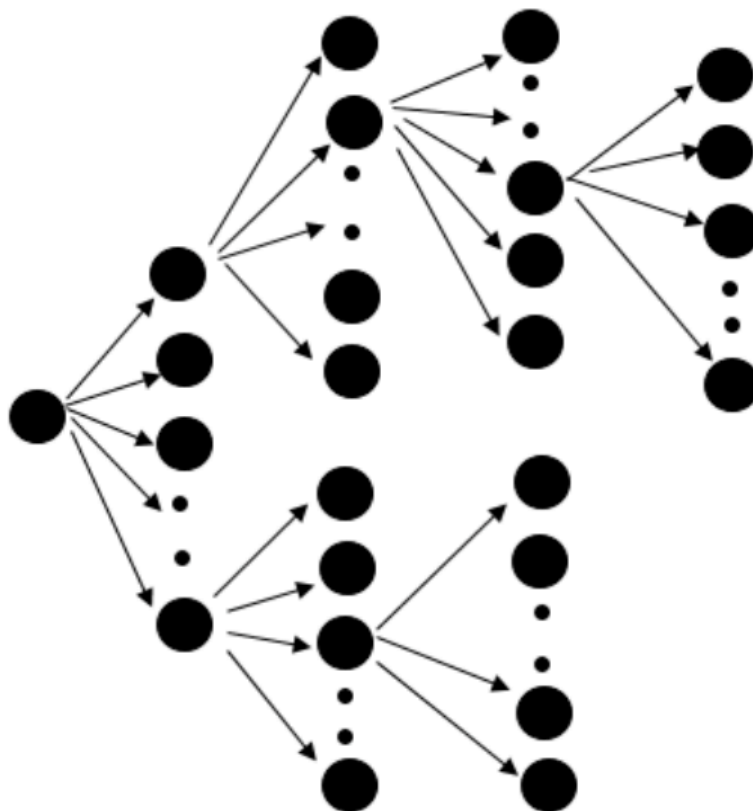
Scenariusz 1 i 3 (Rys. 9. ) okazały się najlepsze w ratowaniu ludzkiego życia poprzez skutkujące mniejszą dynamiką rozprzestrzeniania się wirusa oraz mniejszą liczbę zakażeń i zgonów. W przypadku braku możliwości wprowadzenia lockdownu scenariusz 7 okazał się najskuteczniejszy w zachowaniu ludzkiego życia.

Wyniki symulacji okazały się zbieżne z publikowanymi w literaturze. Wykazano też nieefektywność stosowanych polityk na przykładzie USA, Szwecji oraz Brazylii w odniesieniu do zachowania ludzkiego życia. Kraje starające się zachować jak najlepszy stan gospodarki poprzez

unikanie obostrzeń izolacji, finalnie poniosły koszt w postaci śmierci ludzi, koszty społeczne mające negatywny wpływ na gospodarkę.

### 1.7.3 Model Bhatnagar'a

Bhatnagar (2020) zaproponował ABM umożliwiający predykcję nowych przypadków zakażeń z uwzględnieniem lockdown'u. Bhatnagar wskazuje, że model może również prognozować inne statystyki jak liczba zgonów lub ozdowień zachowując przy tym dostateczną precyzję. Model wykazał względnie wysoką precyzję dla danych dostępnych z różnych krajów. Model jest w stanie wykazać, czy kraj wszedł w III etap pandemii (Euro WHO, (2022)), czyli taki, w którym patogeny przechodzą tylko kilka cykli wtórnej transmisji między ludźmi. Model potwierdził relatywnie szybsze rozprzestrzenianie się choroby we wczesnych miesiącach pandemii we Włoszech niż we Francji.



Rys. 10. Struktura drzewiasta obrazująca scenariusz rozprzestrzeniania się COVID-19 w zaproponowanym przez Bhatnagar (2020) modelu predykcji. Źródło: Bhatnagar (2020)

W założeniach tego modelu przyjęto, że zarażona osoba przybywa do kraju i dochodzi do kontaktu z innymi osobami, co skutkuje rozprzestrzenianiem choroby zakaźnej. Scenariusz zobrazowano w postaci drzewiastej na Rys. 10. Osoba zarażona jest oznaczona jako węzeł  $A$ . Konsekwencją braku identyfikacji aktywnego węzła  $A$  jest przyjęcie założenia, że w kontakcie będą wszystkie inne  $a$  węzły (osoby) w pojedynczej jednostce czasu (np. dzień, godzina, minuta). Węzły  $a$  stają się aktywnymi węzłami  $A$  i mogą zarażać poprzez kontakt kolejne osoby  $a$ . Model się zapętla i przyjmuje postać matematyczną:

$$T_{cases} = 1 + a + (1 + a)a + (1 + a + (1 + a)a)a + (1 + a + (1 + a)a + (1 + a + (1 + a)a)a)a + \dots \quad (7)$$

gdzie  $T_{cases}$  oznacza sumaryczną liczbę zakażeń.

Po przekształceniach algebraicznych oraz założeniu, że  $N$  to numer okresów w szeregu odpowiadający  $N$  odstępom czasu w mierzonych jednostkach, otrzymujemy równanie:

$$T_{cases} = (1 + a)^{N-1} \quad (8)$$

Zakładając, że jedna osoba w jednej jednostce czasu  $a = 1$  może zarazić tylko pojedynczą osobę  $a = 1$ , to cała populacja  $P$  danego kraju będzie zarażona, jeśli zostanie spełniona nierówność:

$$P \leq 2^{N_p-1} \quad (9)$$

gdzie  $N_p$  to najmniejsza wartość  $N$ , która osiąga tę nierówność.

Na podstawie powyższego modelu Bhatnagar oszacował, że licząca 1,3 miliarda populacja Indii w środowisku bez jakichkolwiek ograniczeń zostanie zarażona w ciągu 31 dni, natomiast populacja Włoch licząca 60 milionów ludzi: w przeciągu 26 dni. Oznacza to, że w wielkość populacji nie objętej ograniczeniami nie ma większego wpływu na szybkość rozprzestrzeniania się choroby COVID-19.

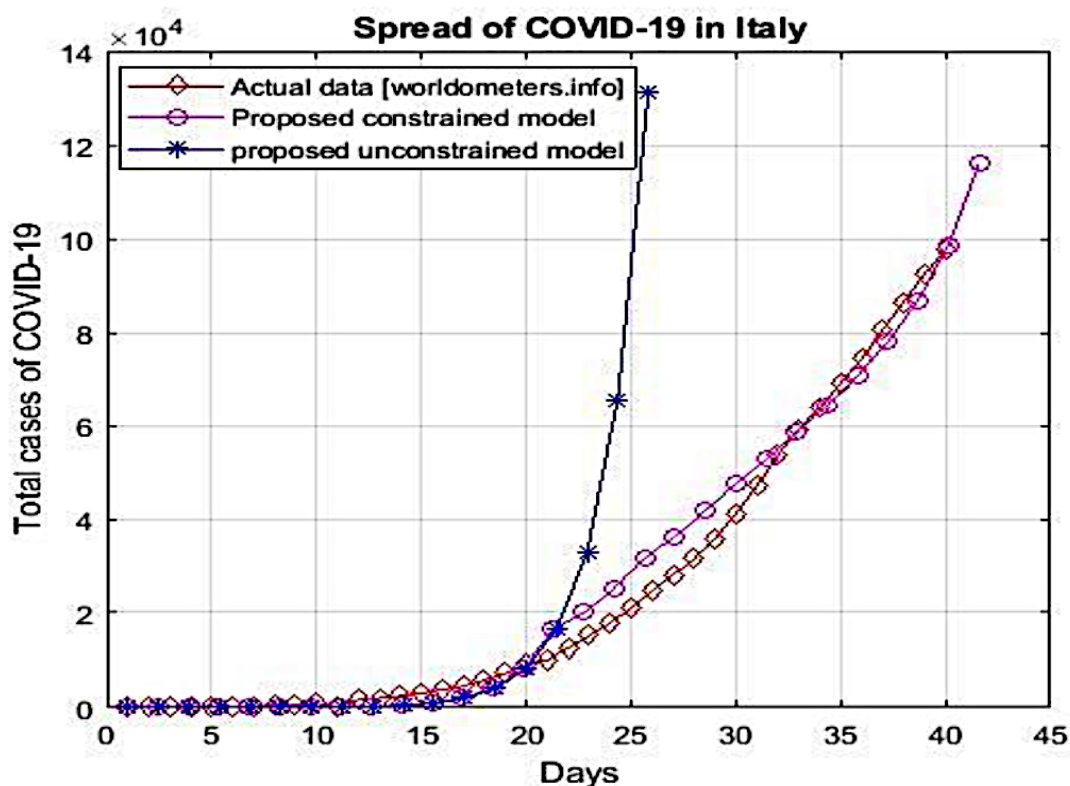
Bhatnagar rozbudował model o kolejne założenia i ograniczenia. Wprowadził czas inkubacji (aktywności) wirusa poprzez  $k$  dni, czyli zakażone węzły i efekty ich działania po upływie  $k$  dni są nieaktywne. Autor wprowadził również wskaźnik  $\frac{\text{rozprzestrzenianie}}{\text{infekcja}}$ . Zakładając jego wartość na poziomie np. 1,46 przyjmujemy zakażenie przez jeden węzeł kolejnego węzła w ciągu 1,46 dnia. Kolejnym założeniem jest  $a = 1$ , czyli zarażanie tylko 1 osobnika oraz  $k = 14$ , oznaczające wyzdrowienie osobnika po 14 dniach. Wpływ ograniczenia mobilności poprzez lockdown,

skutkujący spowolnieniem rozprzestrzeniania się wirusa - odzwierciedlono poprzez wprowadzenie wskaźnika korygującego  $C$ . Wobec powyższego, nierówność wyrażona jest jako:

$$T_{cases} = \begin{cases} X_n, & \text{dla } n \leq k, \\ X_n - X_{n-k+1} - C, & \text{dla } n > k \end{cases} \quad (10)$$

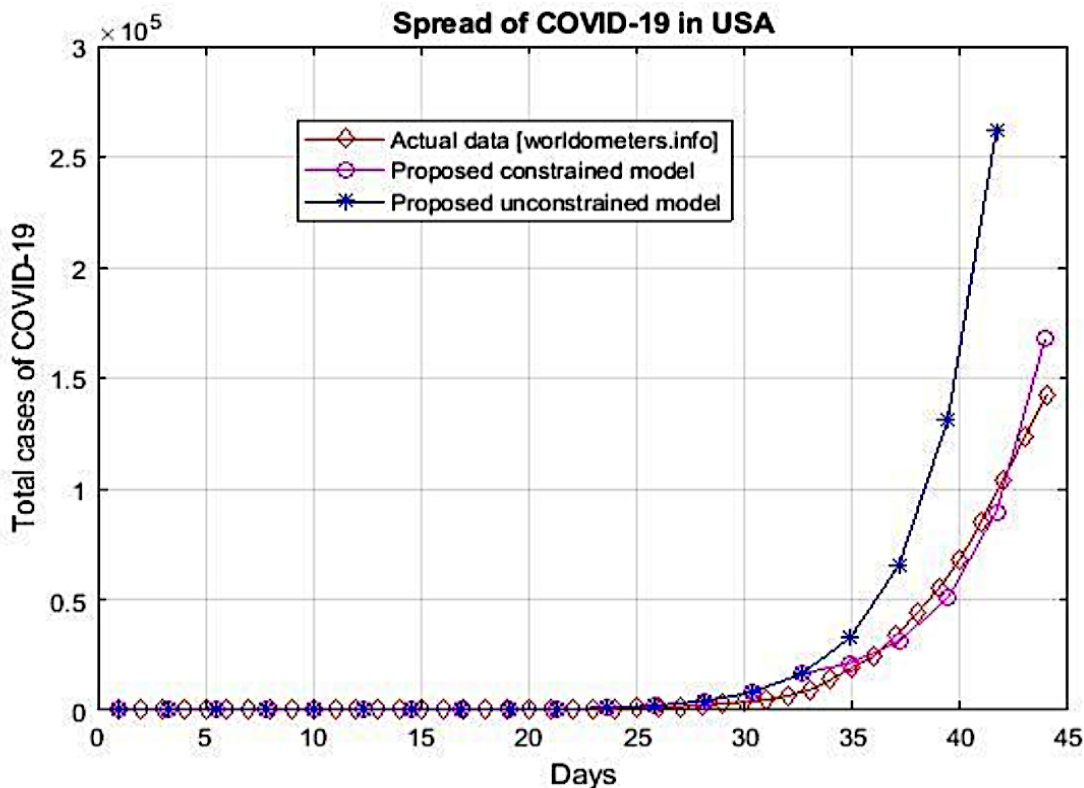
gdzie  $X_n$ ,  $X_{n-k+1}$  są  $n$ -tym i  $n - k + 1$ -tym okresem szeregu czasowego.

Zestawienie symulacji z wykorzystaniem modeli z ograniczeniami oraz bez ograniczeń nałożonych na dane rzeczywiste dla Włoch w początkowej fazie pandemii (do 29 marca 2020 roku) przedstawiono na Rys. 11. Wartości ograniczeń przyjęte do symulacji wyniosły:  $C = 10692$ ,  $k = 14$ ,  $a = 1, \frac{\text{rozprzestrzenianie}}{\text{infekcja}} = 1,46$



Rys. 11. Rozprzestrzenianie się COVID-19 we Włoszech – sumaryczna dzienna liczba zakażeń. Kolorem niebieskim oznaczono prognozę w modelu bez ograniczeń, kolorem różowym prognozę w modelu z ograniczeniami, natomiast kolorem brązowym dane aktualne na podstawie danych z portalu worldometer.info. Źródło: Bhatnagar (2020)

Podobne porównanie symulacji Bhatnagar wykonał dla USA (Rys. 12. ), zmieniając założenia do wartości ograniczeń. Przyjęto  $a = 1, C = 11500, k = 14, \frac{\text{rozprzestrzenianie}}{\text{infekcja}} = 2,26$ . Widoczna jest zbieżność symulacji na podstawie modelu z ograniczeniami oraz rzeczywistych przebiegów. Zastosowanie określonych empirycznie wartości ograniczeń w modelu z ograniczeniami daje możliwość do jego wykorzystania w celach krótkoterminowej predykcji sumarycznej liczby zakażeń w danej populacji w początkowej fazie pandemii.



Rys. 12. Rozprzestrzenianie się COVID-19 w USA – sumaryczna dzienna liczba zakażeń. Kolorem niebieskim oznaczono prognozę w modelu bez ograniczeń, kolorem różowym prognozę w modelu z ograniczeniami, natomiast kolorem brązowym dane aktualne na podstawie danych z portalu worldometer.info. Źródło: Bhatnagar (2020)

#### 1.7.4 Model ARIMA

ARIMA jest metodą zaproponowaną przez Box et al. (2015) i służy do analizy statystycznej z wykorzystaniem szeregów czasowych. Metoda ma zastosowanie dla szeregów stacjonarnych i niestacjonarnych, które można sprowadzić do postaci stacjonarnej. Zastosowanie ARIMA umożliwia predykcję szeregu czasowego z wykorzystaniem różnic pomiędzy wartościami w

szeregu czasowym. ARIMA ma szerokie zastosowanie, np. w prognozie zapotrzebowania danych zasobów bazując na danych historycznych, prognozowanie sprzedaży lub szukanie sezonowości w sprzedaży, szukanie trendów, estymowanie wpływu kampanii marketingowych, sprzedaży nowych produktów itp. ARIMA składa się z 3 komponentów:

1. Autoregresji (AR)
2. Integracji (I)
3. Średniej ruchomej (MA)

Autoregresja to proces, w którym każda wartość jest liniową kombinacją poprzednich wartości.

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_3 y_{t-3} + \dots + \varphi_i y_{t-i} + \varepsilon_t \quad (11)$$

gdzie:

$y_t$  - wartość szeregu w chwili  $t$

$y_{t-i}$  - wartość szeregu w chwili  $t - i$

$\varepsilon_t$  - składnik losowy, zaburzenie w chwili  $t$

$\varphi_i$  - szacowany parametr w celu oceny wpływu  $i$ -tej poprzedniej wartości procesu na wartość aktualną

Rząd autoregresji  $p$  określa liczbę wcześniejszych wartości szeregu czasowego mających wpływ na bieżącą wartość.

Kolejnym istotnym parametrem w metodzie ARIMA jest parametr  $d$  oznaczający stopień różnicowania szeregu. Różnicowanie szeregu (ang. differencing) wykonuje się w celu osiągnięcia stacjonarności procesu i polega ono na obliczaniu pierwszych lub dalszych różnic między kolejnymi wartościami. Pierwsze różnice oblicza się w następujący sposób:

$$\Delta y_t = y_t - y_{t-1} \quad (12)$$

Średnia ruchoma (MA) jest średnią ważoną z ostatnich zaburzeń i składników losowych modelu. W procesie tym zakłada się, że wartość szeregu zależy od zaburzeń w chwili obecnej oraz wcześniejszych. Liczba branych pod uwagę wcześniejszych zaburzeń określana jest poprzez rząd procesu MA - parametr  $q$ .

$$y_t = \beta_1 \varepsilon_{(t-1)} + \beta_2 \varepsilon_{(t-2)} + \dots + \beta_i \varepsilon_{(t-i)} + \varepsilon_t \quad (13)$$

gdzie:

$y_t$  - wartość szeregu w chwili  $t$

$\beta_i$  – wpływ starszych  $i$ -tych zaburzeń szeregu na bieżącą wartość

$\varepsilon_{t-i}$  – składnik losowy będący zaburzeniem w chwili poprzedzającej  $t - i$

$\varepsilon_t$  – wskaźnik losowy będący zaburzeniem w chwili obecnej  $t$

Model ARIMA przy zastosowaniu odpowiednich parametrów będzie równoważny z innymi modelami:

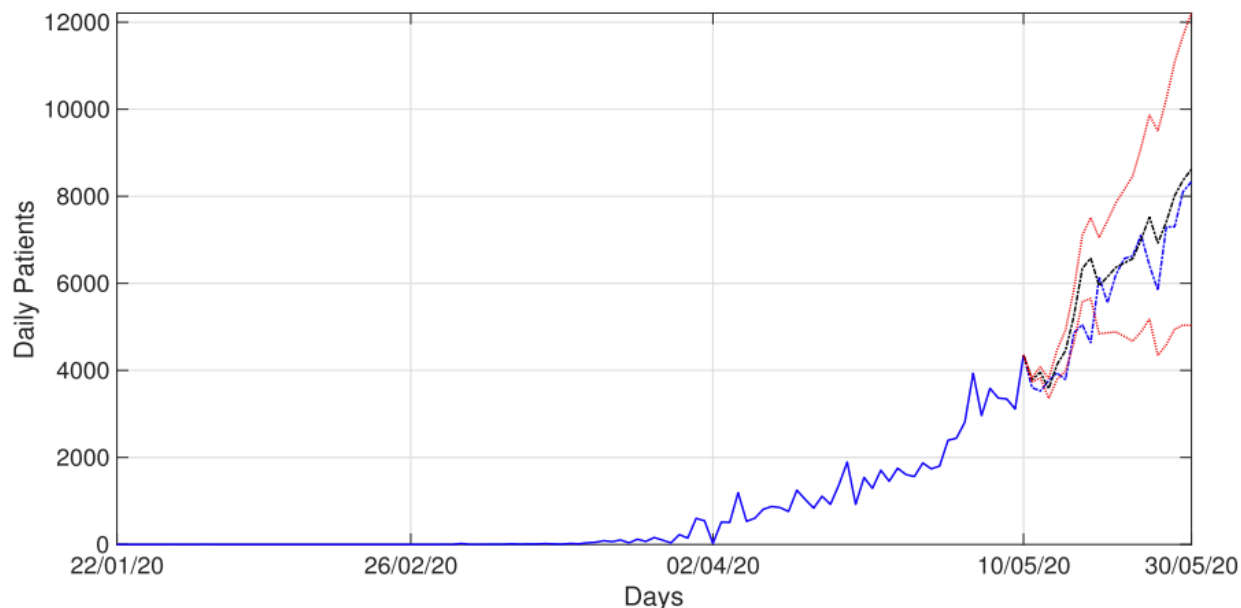
- ARMA – przy braku stacjonarnych danych i wartości parametru  $d = 0$
- AR – model autoregresyjny, brak średnich kroczących lub dane stacjonarne i wartościach parametrów  $p = 0, q = 0$
- MA – model średniej kroczącej bez autoregresji lub danych stacjonarnych i wartościach parametrów  $p = 0, d = 0$ .

Budowanie modelu można podzielić na etapy:

1. Podział zbioru danych na podzbiory: do trenowania modelu (zwykle przyjmuje się 80% wolumenu danych), ewaluacji modelu (20% wolumenu danych).
2. Ustalenie maksymalnych wartości parametrów  $p, d, q$ .
3. Stworzenie zbioru wektorów PDQ będących wynikiem permutacji parametrów  $p, d, q$  w ustalonym zakresie maksymalnym dla każdego z parametrów.
4. Wytrenowanie modelu z wykorzystaniem każdego z osobna wektora PDQ, walidacja modelu. Ostateczny wybór modelu (wektora PDQ zawierającego parametry  $p, d, q$ ) zwykle jest dokonywany na podstawie MSE, AIC lub BIC. Najczęściej przyjmowanym kryterium jest AIC.

Kufel (2020) badał użyteczność modelu ARIMA w predykcji dynamiki COVID-19 w różnych etapach rozwoju pandemii. Model służył do estymacji oraz prognozy rozwoju choroby od jej początkowej fazy, przez okres szczytu zakażeń aż do wyginięcia wirusa.

Sharma et al. (2020) badał możliwości predykcji zakażeń na COVID-19 wykorzystując wstępnie EVDHM do różnicowania szeregów czasowych poprzez ich podział na sekwencje i badanie stacjonarności metodą PPT uzyskanych w ten sposób komponentów. W przypadku spełnienia warunku stacjonarności wykorzystywana jest ARIMA do krótkoterminowych prognoz dla Indii, USA i Brazylii.



Rys. 13. Prognoza szeregu czasowego zakażeń COVID-19 w Indiach za pomocą metody opartej na ARIMA zaproponowanej przez Sharma et al. (2020). Linia ciągła koloru niebieskiego przedstawia dane wykorzystane do trenowania modelu, linia przerywana koloru czarnego oznacza dane prognozowane, linia niebieska przerywana oznacza dane aktualne, natomiast czerwone linie oznaczają RMSE z 95% przedziałem pewności. Źródło: (Sharma et al. (2020)).

Wyniki dla badania prognozy szeregu czasowego dla Indii zamieszczone na Rys. 13. okazały się bardzo obiecująco. Prognoza zawarta w 95% RMSE pokrywała się z aktualnymi danymi w horyzoncie predykcji wynoszącym 109 dni.

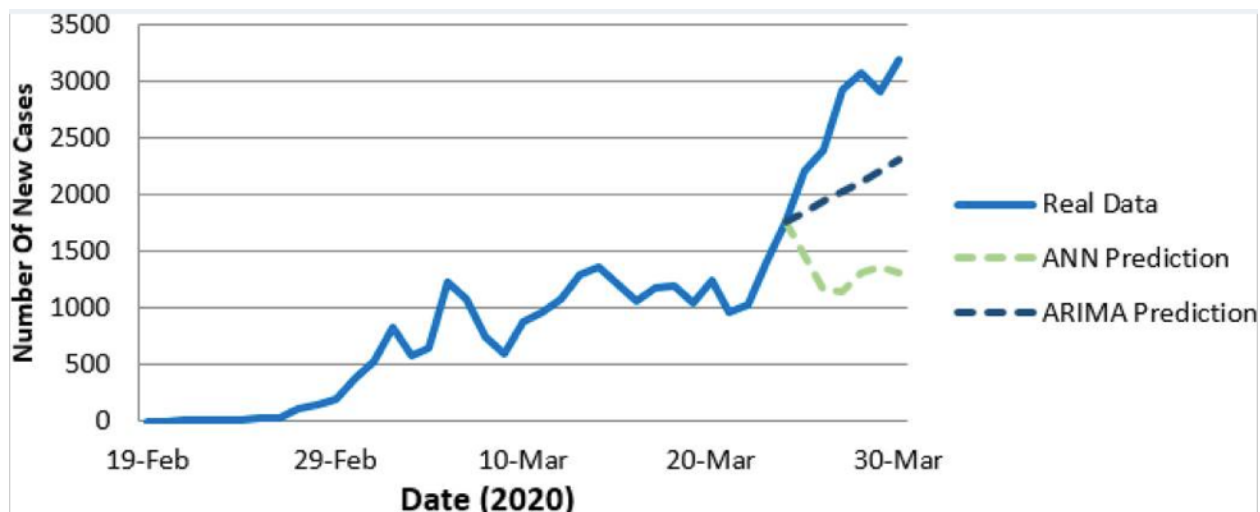
Singh et al. (2020) zidentyfikował 15 krajów o największej licznie zakażeń na świecie w badanym okresie i przeprowadził dla nich predykcję rozprzestrzeniania się wirusa z wykorzystaniem ARIMA. Horyzont czasowy uwzględniony w tej metodzie wynosił 2 miesiące.

Hernandez-Matamoros et al. (2020) zaproponował algorytm do kalkulacji parametrów ARIMA dla 145 krajów, które to parametry dają niski RMSE. Wyniki zostały przetestowane z 10% aktualnych danych. Celem badania jest wykazanie relacji pomiędzy błędem predykcji i innymi zmiennymi, co udało się wykazać dla wielkości populacji. Badanie wykazało, że istnieje relacja pomiędzy rozprzestrzenianiem się wirusa a innymi zmiennymi charakterystycznymi dla krajów należących do tego samego regionu geograficznego. Wykazano również relację pomiędzy populacją kraju i RMSE w predykcji.

García-Cremades et al. (2021) dokonał badania predykcji 14-dniowej w Hiszpanii. Wykorzystanie tylko danych historycznych obarczało prognozy błędem związanym z nieregularnymi komponentami. W predykcji jako nieregularny komponent uwzględniono zatem wpływ mobilności populacji na podstawie danych udostępnionych przez Google LLC "Google COVID-19 Community Mobility Reports" (<https://www.google.com/covid19/mobility/> (2022)). Powyższe podejście poprawiło znacznie precyzję predykcji. W dalszych badaniach kluczowe będzie zidentyfikowanie innych nieregularnych komponentów, np. takich jak szczepienia.

Ala'raj et al. (2021) badał wykorzystanie danych publicznych do zbadania właściwości związanych z pandemią COVID-19 w celu opracowania dynamicznego modelu hybrydowego opartego na SEIRD i współczynnika potwierdzalności z automatycznie dobranymi parametrami. Model składał się ze zmodyfikowanego modelu dynamicznego SEIRD oraz modeli ARIMA. Porównano model SEIRD z wartościami historycznymi populacji zakażonych, wyzdrowiałych i zmarłych podzielonymi przez współczynnik zakażeń, który z kolei jest również parametrem modelu. Pozostałości z pierwszego modelu populacji zakażonych, wyleczonych i zmarłych były korygowane za pomocą modeli ARIMA eliminujących rezydua dynamicznych modeli i poprawiały jakość predykcji. Otrzymane wyniki były testowane na danych rzeczywistych z USA i umożliwiły dokonanie precyzyjnych predykcji do 2 miesięcy.

Moftakhar et al. (2020) badał predykcję COVID-19 w Iranie za pomocą ANN i ARIMA pod względem MSE i MAE. Iloraz współczynników błędu ANN w stosunku do ARIMA wynosił 4,25 dla MSE i 2,11 dla MAE. Z porównania obu metod wynika, że w ANN występuje przeważające niedoszacowanie prognozowanych wartości przypadków COVID-19. Innymi słowy, ARIMA wykazuje się znacznie większą precyzją prognozy zakażeń COVID-19 niż ANN, co przedstawiono na Rys. 14.



Rys. 14. Zaobserwowana i przewidywana liczba nowych przypadków COVID-19 w Iranie z podziałem szeregu czasowego na dane do trenowania modelu (do 24 marca 2020 r.) i do ewaluacji modeli ANN i ARIMA (od 25 marca 2020 r. do 31 marca 2020 r.). Niebieska ciągła linia oznacza dane rzeczywiste, zielona przerywana linia to predykcja za pomocą ANN, natomiast granatowa przerywana linia to predykcja za pomocą ARIMA. Źródło: Moftakhar et al. (2020)

### 1.7.5 Deep learning

Wykorzystanie współczesnych technologii opartych na AI i ML radykalnie poprawiło predykcję, śledzenie kontaktów, prace nad lekiem i szczepionką (Lalmuanawma et al. (2020)).

Zeroual et al. (2020) porównał modele oparte na NN poprzez badanie ich dokładności w 17-dniowej predykcji dziennej liczby zgonów i ozdowień COVID-19 w 6 krajach bazując na danych historycznych ze 148 dni począwszy od stycznia 2020 r. Kryteriami porównawczymi były RMSE, MAE, MAPE, EV i RMSLE. Wśród rozważanych modeli były:

- RNN
- LSTM
- Bi-directional LSTM (BiLSTM)
- GRU
- VAE

Spośród badanych modeli najlepsze wyniki z uwzględnieniem przyjętych kryteriów uzyskano metodą VAE (Tab. 2. ).

Tab. 2. Porównanie dokładności predykcji ozdowień COVID-19 za pomocą wybranych modeli dla populacji Włoch. Źródło: Zeroual et al. (2020).

Country	Model	RMSE	MAE	MAPE	EV	RMSLE
Italy	RNN	4,006,731	3,950,381	23,734	0103	00754
	GRU	3,892,321	382,867	22,989	0017	00705
	LSTM	3,887,027	3,822,325	22,949	0002	00703
	BiLSTM	3,886,682	3,821,869	22,946	0	00702
	VAE	2,273,616	2,250,268	13,537	0789	00163

Wśród metod DL wykorzystywanych do szukania tymczasowych korelacji w predykcji szeregów czasowych są RNN. RNN sprawdza się w przypadku, kiedy poprzednia-wcześniejsza informacja z szeregu czasowego jest potrzebna do zdecydowania o stanie obecnym. Sieci zawierają ukryte stany wpływające na precyzję predykcji, która jest większa niż w innych metodach z wykładniczym wygładzaniem (Zhao et al. (2020)).

RNN i rozszerzenia LSTM są wykorzystywane szeroko w kontekście COVID-19. Pathan et al. (2020) badał za pomocą modelu LSTM predykcję wskaźnika mutacji wirusa SARS-CoV-2 w ludzkim ciele. Nemes i Kiss (2020) badali za pomocą RNN nastroje społeczne związane z COVID-19 na podstawie NLP w mediach społecznościowych.

Dutta et al. (2020) badał predykcję zakażeń, zgonów i ozdowień COVID-19 za pomocą hybrydowego modelu LSTM-GRU. Kryterium porównawczym był RMSE, które najniższą wartość 4,16% uzyskało dla predykcji zgonów, natomiast najwyższą wartość 49,4 dla ozdowień.

Shastri et al. (2020) porównał modele DL w prognozowaniu szeregów czasowych COVID-19 (zakażeń oraz zgonów) na przykładzie Indii i USA (Tab. 3. ). Z rozważanych metod było rozszerzenie RNN w postaci komórek LSTM i ich wariantów. Kryterium porównawczym był MAPE. Rozpatrywanymi wariantami były:

- Stacked LSTM
- Bi-directional LSTM
- Convolutional LSTM

Najniższy błąd MAPE (mieszczący się w przedziale 2-3,3%) z wyżej wymienionych modeli zaobserwowano w modelu Convolutional LSTM, natomiast model Stacked LSTM został oceniony najslabiej. Szczegółowe wyniki porównania MAPE dla wyżej wymienionych modeli przedstawiono w Tab. 3.

*Tab. 3. Porównanie MAPE dla wybranych modeli wariantów LSTM w predykcji zakażeń i zgonów w Indiach i USA. Źródło: Shastri et al. (2020).*

Models	India confirmed cases	USA confirmed cases	India death cases	USA death cases
	MAPE	MAPE	MAPE	MAPE
Stacked LSTM	4.00	10.00	6.66	8.00
Bi-directional LSTM	3.33	6.66	4.00	6.00
ConvLSTM	2.17	2.00	3.33	2.50

## 1.8 Inflacja jako drugi społecznie istotny cel modelowania

W kontekście szybko ewoluującego środowiska gospodarczego, precyzyjne prognozowanie inflacji nabiera istotnego znaczenia, stanowiąc wyzwanie intelektualne oraz praktyczne dla naukowców z dziedziny informatyki i ekonomii. Zastosowanie nowoczesnych metod analitycznych i algorytmicznych do przewidywania zmian inflacyjnych stanowi pole badań, łączące w sobie zaawansowane techniki obliczeniowe z głębokim zrozumieniem zjawisk ekonomicznych. Wśród rozważanych metodologii w przeglądzie literatury szczególną uwagę zwrócono na podejścia ekonometryczne, makroekonomiczne, techniki uczenia maszynowego, statystykę Bayesowską, sieci neuronowe oraz analizę spektralną. Techniki ekonometryczne, takie jak regresja liniowa z optymalizacją algorytmami genetycznymi, modele ARIMA i Holt-Winters, a także Boxa-Jenkinsa i FTSA, demonstrują potencjał informatyki w ekstrakcji wiedzy z danych historycznych. Analiza makroekonomiczna, reprezentowana przez krzywą Philipsa, podkreśla złożoność dynamiki między bezrobociem a inflacją, wyzwalając potrzebę zastosowania zaawansowanych metod analizy. Uczenie maszynowe i sieci neuronowe, z przykładami takimi jak SVR i HRNN, ujawniają nowe możliwości w modelowaniu nieregularnych zestawów danych,

oferując perspektywę na zrozumienie nielinearności i złożoności ekonomicznych cykli inflacyjnych. Statystyka Bayesowska, ze szczególnym wskazaniem na metodę BMA, dostarcza ramy do integracji wiedzy a priori z obserwacjami, umożliwiając bardziej elastyczne modelowanie przyszłych trendów inflacyjnych. Dodatkowo, zaawansowane metody ekonometryczne, w tym regresja MIDAS i estymator SLS, jak również modele nieparametryczne takie jak regresja splajnowa wielowymiarowa i B-spline, prezentują innowacyjne podejścia do analizy ekonomicznej, wykorzystując zaawansowane narzędzia statystyczne i obliczeniowe do modelowania i prognozowania inflacji.

W omawianym podrozdziale zaprezentowano analizę danych dotyczących inflacji, reprezentowanej przez wskaźnik CPI, dla państw należących do Organizacji Współpracy Gospodarczej i Rozwoju (OECD). Wykorzystując technikę wizualizacji danych jaką są wykresy pudełkowe, dokonano ilustracji rozkładu wartości inflacji zarówno w perspektywie całej dostępnej historii danych, jak i w wycinku obejmującym 100 najnowszych miesięcy. Szczególna uwaga została poświęcona analizie szeregów czasowych dla wyselekcjonowanych krajów, które charakteryzują się unikalnymi wartościami inflacji.

Źródła danych zostały dokładnie zidentyfikowane, a następnie opisano proces ich przygotowania do szczegółowej analizy za pomocą oprogramowania MATLAB, co podkreśla znaczenie odpowiedniej obróbki danych w celu zapewnienia ich wiarygodności i użyteczności w analizie. Punktem kulminacyjnym podrozdziału jest prezentacja i interpretacja szeregu czasowego inflacji dla Polski w okresie od roku 2006 do 2023, co pozwala na dogłębne zrozumienie trendów inflacyjnych w kontekście specyfiki polskiej gospodarki.

### 1.8.1 Modele inflacji i metody jej obliczania

Zrozumienie i przewidywanie (predykcja szeregów czasowych) inflacji jest kluczowym wyzwaniem w ekonomii i polityce monetarnej. W ostatnich latach, badania w tej dziedzinie znacząco ewoluowały, obejmując szeroki zakres metodologii, od tradycyjnych modeli ekonometrycznych po zaawansowane techniki uczenia maszynowego i innowacyjne podejścia w analizie szeregów czasowych. Te różnorodne podejścia odzwierciedlają złożoność inflacji jako wskaźnika ekonomicznego i podkreślają potrzebę holistycznego ujęcia w jej modelowaniu.

W kontekście niniejszej rozprawy, istotne jest zrozumienie, w jaki sposób różne modele i techniki mogą być wykorzystywane do prognozowania. Poniżej dokonano porównania kluczowych prac w dziedzinie prognozowania inflacji, z uwzględnieniem ich metodologicznych fundamentów, zakresu zastosowania oraz ograniczeń.

Dharma et al. (2020) badali prognozowanie inflacji metodą regresji liniowej. Znalezienie odpowiedniego modelu potraktowano jako zadanie optymalizacyjne z wykorzystaniem optymalizacji algorytmami genetycznymi. Wielokrotna regresja liniowa (MLR) była używana w tym badaniu, mając na celu znalezienie wzorca, który jest najbliższy charakterystyce historycznych stóp inflacji w Indonezji, gdzie wzorec ten jest reprezentowany w postaci funkcji liniowych. Dobierając różne parametry algorytmu genetycznego osiągnięto dokładność predykcji inflacji na poziomie  $MSE=0.1099$ .

Stock & Watson (2008) badali predykcję inflacji opartą na krzywej Philipsa w Stanach Zjednoczonych. Krzywa Philipsa opisuje związek między stopą bezrobocia a inflacją. Zakłada się, że istnieje odwrotny związek między tymi zmiennymi: kiedy stopa bezrobocia jest niska, inflacja jest wysoka i odwrotnie. Stock & Watson stwierdzili epizodyczną dokładność prognozowania inflacji z wykorzystaniem krzywej Philipsa. Oznacza to, że dawała ona zadowalające wyniki dla wybranych spośród badanych okresów.

Epizodyczną dokładność prognozowania inflacji na przykładzie strefy Euro potwierdzają również Bańbura & Bobeica (2023). Autorzy zwracają uwagę na istotny wpływ uwzględniania trendu na dokładność prognozy.

Atkeson & Ohanian (2001) wskazują, że krzywe Philipsa były kluczowym elementem modeli makroekonomicznych przez ostatnie 40 lat. Ich przegląd dowodów wskazuje, że nie opracowano wersji krzywej Philipsa, która skutecznie prognozowałaby inflację lepiej niż proste modele. W związku z tym uważają oni, że poszukiwanie kolejnego modelu prognozującego opartego na krzywej Philipsa powinno zostać porzucone. Badania wskazują również, że żaden z potencjalnych wskaźników inflacji nie jest szczególnie skuteczny.

Ülke et al (2018) badali oni porównanie jednoziennej i wieloziennej analizy szeregów czasowych oraz modeli uczenia maszynowego w kontekście prognozowania inflacji (wersje podstawowe i niestandardowe wskaźników CPI i PCE). Ich empiryczne wyniki wskazują, że:

- a) SVR (Maszyna Wektorów Nośnych w Regresji) daje lepsze wyniki niż inne modele w prognozowaniu inflacji podstawowej według wskaźnika PCE.

- b) ARDL zapewnia najniższy błąd prognozy i najwyższą dokładność prognozy inflacji podstawowej według wskaźnika CPI.
- c) Modele uczenia maszynowego działają lepiej z bardziej zmiennymi i nieregularnymi szeregami.

Na podstawie tych obserwacji doszli do wniosku, że nie ma jednego najlepszego modelu do prognozowania inflacji.

Wright (2009) zwraca uwagę na metodę BMA w kontekście prognozowania inflacji w USA. Stwierdził, że BMA zwykle przewyższa tradycyjne metody prognozowania oparte na równomiernym ważeniu. Wynik ten jest spójny dla różnych miar inflacji. Ważnym atutem BMA jest jego uniwersalność - można go stosować do różnych miar inflacji i horyzontów prognoz.

Maldeni & Mascenghe (2021) badali prognozowanie inflacji w Sri Lance opartej na zdefiniowanym indeksie CCPI przy użyciu metod uczenia maszynowego. Badanie obejmowało porównanie sześciu modeli ML: RF, ANN, XGBoost, SVR, KNN i LR. Najbardziej efektywne modele to SVR, RF i KNN. Chociaż ANN i RF są popularne w literaturze, wielkość zbioru danych miała wpływ na ich prognozy. SVR okazał się szczególnie dokładny dla mniejszych zbiorów danych.

Barkan et al (2023) proponują hierarchiczną architekturę opartą na rekurencyjnych sieciach neuronowych (HRNN) do prognozowania rozłożonych składników inflacji CPI, które są kluczowe dla głębszego zrozumienia jej dynamiki. Podkreślają, że dynamika inflacji CPI jest czasami napędzana przez składniki niezwiązane z celami polityki banku centralnego, takie jak ceny żywności i energii. Prognoza CPI na poziomie rozłożonym daje dokładniejszy obraz źródeł i cech przyszłych nacisków inflacyjnych w gospodarce. Model HRNN został zaprojektowany do prognozowania rozłożonych składników CPI, ale autorzy wierzą, że może być on również przydatny do prognozowania innych hierarchicznych szeregów czasowych, takich jak PKB.

Hassani et al. (2013) zastosował techniki analizy jednostkowego spektrum, zarówno w wersji jedno- jak i wielowymiarowej (SSA i MSSA), do prognozowania dynamiki inflacji w USA. W swoich prognozach oparto się na indeksach cen konsumenckich oraz indeksach cen PKB. Dodatkowo, autorzy porównywali efektywność swojego podejścia z innymi metodami prognozowania, takimi jak krzywa Phillipsa, model autoregresyjny AR(p), model dynamicznych czynników oraz błędzenie losowe traktowane jako podstawowa metoda prognozowania. Do prognoz używali zarówno krótkoterminowych okien czasowych (kwartalnych), jak i

długoterminowych (od 1 do 6 lat). Autorzy zastosowali w badaniu techniki SSA i MSSA do prognozowania ekonomicznych szeregów czasowych. Wyniki wykazały, że MSSA znacząco przewyższa inne metody prognozowania inflacji. W porównaniu zastosowano metodę RW, krzywą Phillipsa, model DFM oraz modele AR(p) (Autoregresyjne). W każdym przypadku MSSA okazała się bardziej dokładna w prognozowaniu inflacji niż te tradycyjne metody ekonometryczne.

Shinkarenko et al.(2021) analizował dynamikę szeregów czasowych inflacji na Ukrainie w okresie od stycznia 2010 do września 2020 korzystając z modeli ARIMA\*ARIMAS (Box-Jenkins) i wygładzania wykładniczego Holt-Winters do modelowania i prognozowania inflacji. Badanie wykazało, że model Holt-Winters osiągnął minimalny błąd prognozy. Przewidywania oparte na modelu Holt-Winters wykazały dobrą zgodność z rzeczywistymi wartościami inflacji, wskazując na jego dokładność w krótkoterminowych prognozach.

Zafar et al. (2015) badał prognozowanie inflacji w Pakistanie, wykorzystując metodę Boxa-Jenkinsa oraz analizę szeregów czasowych funkcjonalnych (FTSA). FTSA traktuje dane jako funkcje ciągłe, co pozwala na bardziej złożone analizy trendów i wzorców. Badanie wykazało, że FTSA zapewnia lepszą dokładność prognoz w porównaniu z modelami SARIMA. W pracy zastosowano również metody wykrywania i korygowania outlierów, transformacji danych i estymacji funkcji wygładzających na przekształconych danych.

Fibriyani & Chamidah (2021) porównali podejście modelu regresji nieparametrycznej z wykorzystaniem lokalnego estymatora wielomianowego do podejścia modelu ARIMA w prognozowaniu stopy inflacji w Indonezji. Na podstawie MAPE stwierdzono przewagę pierwszego modelu nad ARIMA.

Breitung & Roling (2015) stosowali regresję MIDAS i estymator SLS, umożliwiające analizę danych o różnych częstotliwościach, co okazało się kluczowe w modelowaniu inflacji, oferując elastyczność i nowoczesne podejścia statystyczne. Ich dokładność, potwierdzona przez symulacje Monte Carlo, podkreśla znaczenie zaawansowanych metod w analizie ekonomicznej.

Prahutama & Santoso (2018) modelowali inflację za pomocą regresji nieparametrycznej B-spline w określonych sektorach (transportu, komunikacji i usług finansowych) dla lepszego zrozumienia wzorców w danych. Badania te skupiały się na modelowaniu inflacji w określonych sektorach gospodarki w oparciu o dane inflacyjne z tych konkretnych sektorów gospodarki. Autorzy optymalizują model za pomocą GCV, czyli Generalized Cross-Validation (uogólniona

walidacja krzyżowa). Autorzy wskazują za Misztal (2006), że funkcja splajnow jest silnie zależna od rozkładu danych, stopnia funkcji, a zwłaszcza od wyboru liczby i położenia węzłów.

Caraka & Sugiyarto (2016) zajmowali się na analizą i modelowaniem wskaźnika inflacji w Indonezji, używając do tego celu regresji parametrycznej i nieparametrycznej, ze szczególnym uwzględnieniem regresji splajnowej wielowymiarowej. Celem było określenie najdokładniejszego modelu prognozowania inflacji. Badanie wykazało, że modele nieparametryczne wykazują wyższą dokładność i lepszy  $R^2$  w porównaniu z modelami parametrycznymi, analizując związki między inflacją a czynnikami takimi jak ceny ropy, podaż pieniądza i kurs rupii wobec dolara.

Binder et al. (2013) porównywał dokładność predykcji metodami splajnowymi i wielomianami ułamkowymi pod względem dostępności danych. Gdy dostępna jest mała ilość danych, metody splajnow są nieco lepsze, zapewniając mniejszą zmienność w wynikach. Przy średniej ilości danych, wielomiany ułamkowe wykazują mniejszy błąd predykcji niż splajny, szczególnie dla prostszych funkcji. Natomiast przy dużej ilości danych, różnice w dokładności między splajnami a MFP stają się minimalne, prowadząc do podobnych modeli.

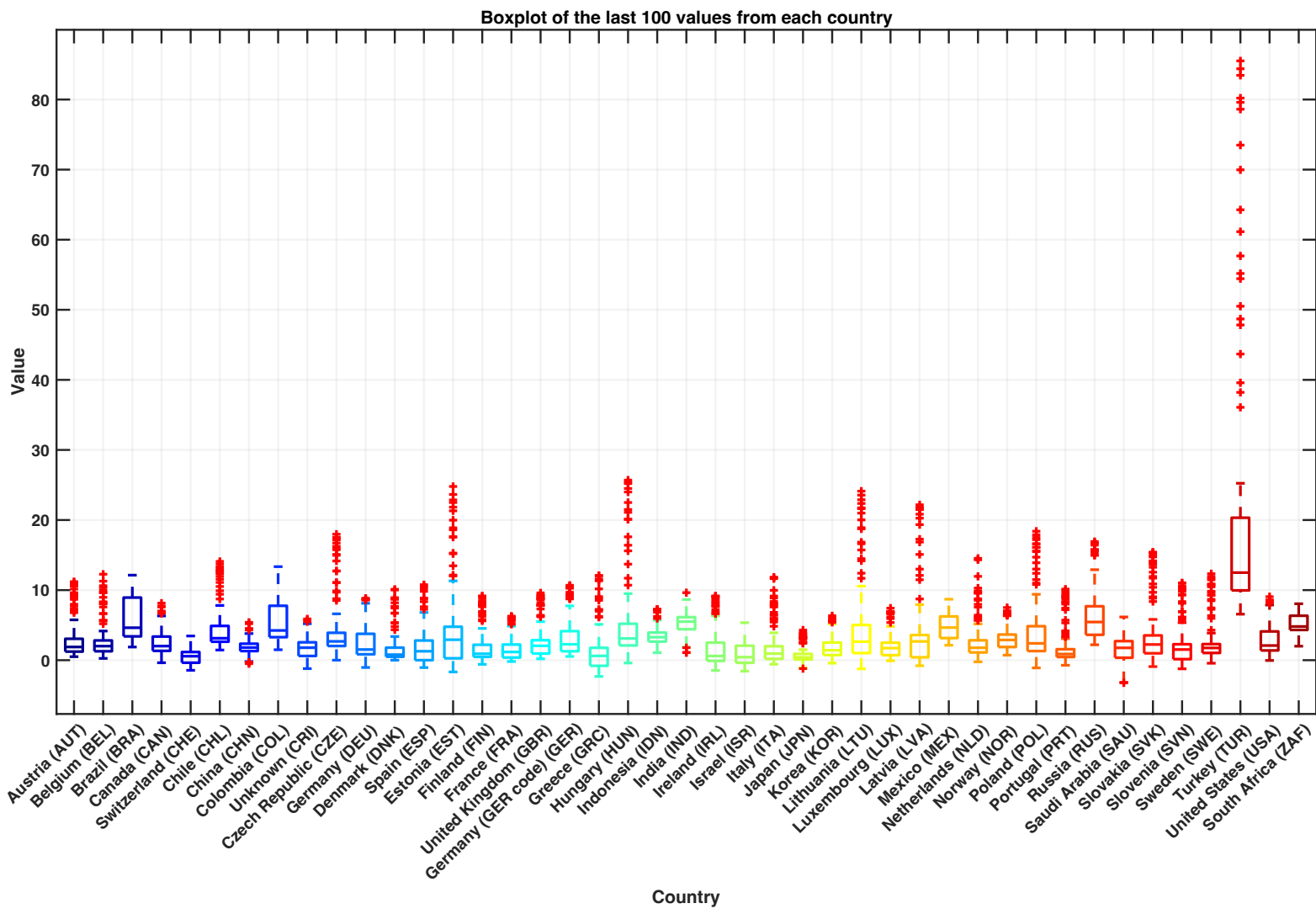
### 1.8.2 Inflacja w państwach OECD.

Inflacja w przyjętym podejściu bazuje na wskaźniku cen konsumpcyjnych (CPI). Jest on miarą zmian cen towarów i usług konsumpcyjnych zakupionych przez gospodarstwa domowe. CPI jest szeroko stosowany jako wskaźnik inflacji, ponieważ odzwierciedla zmiany w kosztach życia, pokazując, jak ceny zmieniają się z czasem. Wzrost CPI oznacza wzrost cen, co jest jednym z głównych wskaźników inflacji. CPI jest bezpośrednio powiązany z inflacją, ponieważ mierzy stopień, w jakim ceny ogólnie rosną, wpływając na siłę nabywczą pieniądza.

Dane badanych krajów zostały pobrane ze strony OECD (<https://data.oecd.org/price/inflation-cpi.htm>). Wybrano dane wszystkich dostępnych państw oraz maksymalny zakres dat obejmujący zgromadzone dane. Pominęto dane zagregowane dla wielu państw, np. grup EU27 – obejmującej kraje Unii Europejskiej, OECD, G7 i G20. W wstępnej analizie danych wykluczono z badania kraje zawierające mniej niż 101 rekordów, ponieważ taki jest minimalny zbiór danych niezbędny do przeprowadzenia badania za pomocą zaimplementowanego algorytmu. Następnie wyodrębniono dane dla każdego kraju osobno i

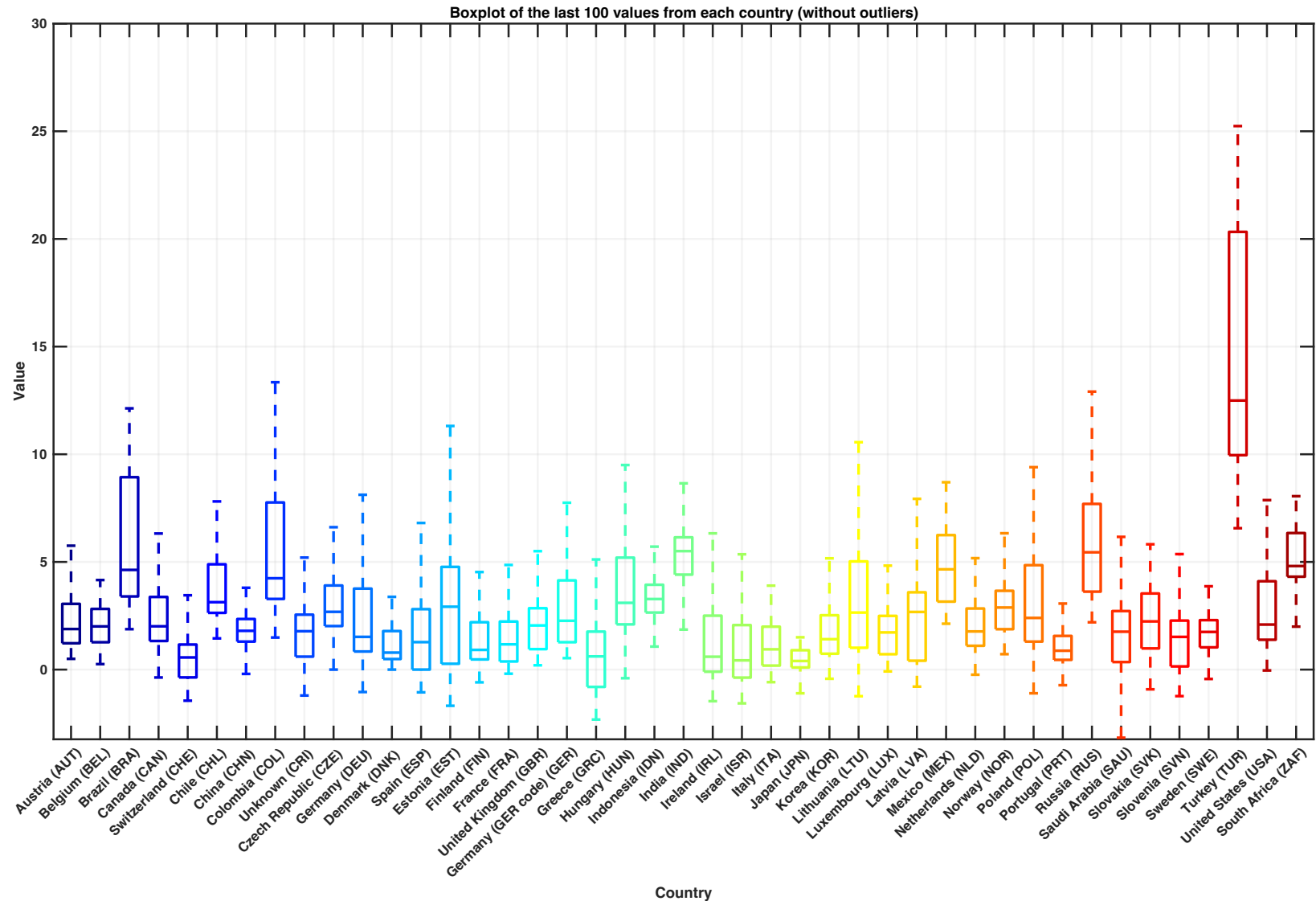
zapisywano je w postaci pliku MATLAB z nazwą w formacie duże CPI\_<kod\_kraju>.m w postaci wektora Y

Krokiem było sprawdzenie uzyskanych danych. W celu analizy danych dla wybranych państw wygenerowano wykres pudełkowy dla 100 ostatnich elementów szeregu czasowego inflacji.



Rys. 15. Wykres pudełkowy wartości inflacji ze 100 ostatnich miesięcy w poszczególnych krajach. Stan na październik 2023.  
 (Źródło: <https://data.oecd.org/price/inflation-cpi.htm>)

Na wygenerowanym wykresie pudełkowym (Rys. 15. ) widoczna jest duża ilość outlierów. Ich obecność prawdopodobnie spowodowana jest wyjątkowymi zdarzeniami gospodarczymi i społecznymi, które miały miejsce w ostatnich latach. Wśród nich znaczącą rolę mogły odegrać globalne kryzysy, takie jak pandemia COVID-19, która wywołała znaczące zakłócenia w łańcuchach dostaw i produkcji, prowadząc do niestabilności cen. Dodatkowo, inne czynniki, takie jak kryzysy społeczno-ekonomiczne, zmiany polityczne, czy naturalne katastrofy, również mogły przyczynić się do anomalii w danych inflacyjnych. Aby uzyskać jaśniejszy obraz rozkładu wartości inflacji, wygenerowano kolejny wykres z pominięciem outlierów (Rys. 16. ). Pozwoli to na dokładniejszą analizę i porównanie danych inflacyjnych w poszczególnych krajach, zwiększając tym samym czytelność i precyzję analizy.



Rys. 16. Wykres pudełkowy wartości inflacji (z pominięciem outlierów) ze 100 ostatnich miesięcy w poszczególnych krajach. Stan na październik 2023. (Źródło: <https://data.oecd.org/price/inflation-cpi.htm>)

Z powyższej analizy wartości inflacji ze 100 ostatnich miesięcy uwagę zwracają szeregi czasowe następujących krajów:

**Turcja** – Turcja miała największy zakres wartości inflacji w ciągu ostatnich 100 miesięcy, z wartościami wahającymi się od 6.57% do 85.51%. Ten ekstremalny zakres wskazuje na znaczne niestabilności gospodarcze. Badanie przyszłej prognozy inflacji dla Turcji może dostarczyć ważnych informacji na temat skutków polityki gospodarczej.

**Estonia** – Estonia prezentuje znaczne wahania inflacji, z wartościami od -1.68% do 24.79%. Kraje z taką zmiennością inflacji mogą być trudne do prognozowania, co czyni je interesującym przedmiotem badań.

**Węgry** – Węgry, podobnie jak Estonia, miały duże wahania inflacji, ale ich średnia i mediana były wyższe. Badanie Węgier może dostarczyć informacji na temat skutków decyzji gospodarczych w regionie Europy Środkowo-Wschodniej.

**Islandia** – Islandia jest krajem, który doświadczył poważnych kryzysów finansowych w przeszłości. Badanie prognoz inflacji dla Islandii może dostarczyć ważnych informacji na temat procesu odbudowy po kryzysie oraz skutków polityk gospodarczych w małym kraju o otwartej gospodarce.

### 1.8.3 Inflacja w Polsce.

Zmiany poziomu inflacji w Polsce przeanalizowano w okresie od 2006 r. do września 2023 r. Dane zostały przedstawione na Rys. 17.



Rys. 17. Poziom inflacji w Polsce w okresie 1.2006 r. – 9.2023 r. (Źródło: opracowanie własne na podstawie <https://data.oecd.org/price/inflation-cpi.htm>).

Analizując dane inflacji w Polsce (Rys. 17. ) w kontekście wydarzeń politycznych i gospodarczych, możemy wyróżnić kilka kluczowych okresów:

- 2006-2008: Ten okres był charakteryzowany przez dynamiczny wzrost gospodarczy w Polsce, częściowo napędzany silnym popytem krajowym i inwestycjami. Polska w tym okresie charakteryzowała się silną gospodarką, napędzaną przez bezpośrednie inwestycje zagraniczne i rosnącą produktywność (Drozdowicz-Bieć (2011)). Kraj doświadczył w tych latach stosunkowo wysokich stóp inflacji, często przypisywanych rosnącemu popytowi krajowemu i wzrostowi cen surowców na świecie.
- 2009-2010: Globalny kryzys finansowy z 2008-2009 miał znaczący wpływ na polską gospodarkę. Jednak Polska była jednym z niewielu krajów europejskich, które uniknęły recesji. Drozdowicz-Bieć (2011) wskazuje, że polska waluta i jej pływający kurs wymiany były istotnym czynnikiem, który przyczynił się do uniknięcia recesji w Polsce. Te aspekty miały pozytywny wpływ na eksport, wzmacniając go, a także obniżając koszty importu. Dzięki temu polski eksport utrzymał się na stabilnym poziomie w okresie recesji. Polityka fiskalna rządu w połączeniu z elastyczną polityką kursu walutowego pomogła złagodzić wpływ kryzysu.
- 2011-2014: po kryzysie finansowym, Polska doświadczała okresu umiarkowanego wzrostu gospodarczego i inflacji. Na poziom inflacji w tych latach wpływały różne czynniki, w tym globalne warunki gospodarcze, krajowy popyt i środki polityki monetarnej.
- 2015-2018: to okres zmian politycznych. Rząd wprowadził programy wydatków socjalnych i reform gospodarczych, które mogły mieć wpływ na inflację. Gospodarka nadal rosła, ale stopy inflacji fluktuowały, odzwierciedlając zarówno zmiany w polityce krajowej, jak i globalne warunki gospodarcze.
- 2019-2020: pandemia przyniosła znaczące wyzwania gospodarcze na całym świecie. Gospodarka Polski skurczyła się w 2020 roku z powodu wpływu pandemii. Rząd i bank centralny wprowadzili środki stymulujące, aby wesprzeć gospodarkę, co wpłynęło na stopy inflacji. Grabia (2020) wskazuje, że w trakcie pandemii ceny przestały być procykliczne, a początkowy spadek inflacji został szybko zastąpiony wzrostem. Narodowy Bank Polski zastosował środki prowadzące do niemal zerowych stóp procentowych, co osłabiło skuteczność tego narzędzia w kryzysie.

- 2021-do teraz: w okresie tym podejmowano wysiłki na rzecz odbudowy i globalną przebudowę warunków gospodarczych, w tym zakłócenia w łańcuchach dostaw i zmiany w cenach energii. Te czynniki, wraz z krajowymi środkami politycznymi, nadal wpływają na stopy inflacji w Polsce.

## 1.9 Sformułowanie problemu badawczego

Model SIR i pochodne sprawdzają się dla krótkotrwałych interwałów, są niedokładne dla długoterminowych prognoz. Modele sztucznej inteligencji zależą od wielu etapów uczenia się, ich ważność jest ograniczona wystarczającą liczbą zestawów danych treningowych. Prognozowanie na poziomie dni bazujące na szeregach czasowych są kolejnym podejściem, które łatwo powieli poprzednie wzorce i nie może przewidzieć trendu zmian. Modelowanie oparte na agentach symuluje współtowarzyszy osobników (agentów) w celu obliczenia rozprzestrzeniania się choroby w społeczności. Niemniej jednak takie modele opierają się na parametrach na poziomie populacji, które nie są jeszcze znane. Istotnym problemem jest mała ilość informacji i brak wiarygodnych danych dla nowopowstałej choroby. Poszukuje się przydatnego podejścia do przewidywania rozprzestrzeniania się wirusa na poziomie kraju. Podejście powinno polegać na opracowaniu metody i modelu umożliwiającego wykonanie precyzyjnych prognoz niezależnie od fazy pandemii, dla dowolnego kraju, bez względu na jego położenie geograficzne, na podstawie publicznie dostępnych danych online w horyzoncie predykcji do 3 tygodni. Proponowana metoda powinna uwzględniać poziom szczepień w danym kraju. Metody powinny zwracać precyzyjne prognozy zwłaszcza po zmianie dynamiki zakażeń, co umożliwi ewaluację wprowadzanych i planowanych decyzji związanych z zapobieganiem rozprzestrzeniania się wirusa.

Tradycyjne metody predykcji inflacji, takie jak modele ekonometryczne i metody bazujące na wskaźnikach makroekonomicznych, często nie radzą sobie z złożonością i dynamicznymi zmianami w gospodarce. W kontekście inflacji problem badawczy koncentruje się na opracowaniu metody, która umożliwi precyzyjne prognozy inflacji dla krajów na poziomie miesięcznym, dostarczając jednomiesięcznych prognoz potrzebnych do podejmowania decyzji ekonomicznych na różnych szczeblach zarządzania. Taka metoda powinna być elastyczna i zdolna do adaptacji do dynamicznych zmian w gospodarce, co pozwala na efektywniejsze reagowanie na złożone interakcje ekonomiczne i nielinerne zależności. Metoda musi być odporna na zewnętrzne szoki,

takie jak kryzysy finansowe lub nagłe zmiany polityczne, być niezawodną w niepewnych warunkach gospodarczych., pozwalając na odpowiednie planowanie i podejmowanie działań w zakresie polityki gospodarczej i inwestycyjnej.

## 1.10 Hipoteza badawcza

Dla wybranego państwa i dowolnie wybranego interwału czasowego daje się opracować model zmian statystycznych w pandemii i inflacji oraz oparta na nim metodę prognozowania, taką, która w porównaniu z metodami referencyjnymi daje błędy najmniejsze.

Dokładność tak wyselekcjonowanej metody powinna być oceniana jako wystarczająca w percepcji społecznej i wg ocen środowiska akademickiego konfrontującego te dokładność z wynikami umieszczanymi w publikacjach naukowych.

Podejście takie zapewnia falsyfikowalność znamionującą naukowość metody, można bowiem wyobrazić sobie sytuację, gdy nawet najlepsza metoda (w sensie uzyskiwanych dokładności) nie zapewni rozwiązania satysfakcjonującego (w aspekcie konfrontacji z wynikami w publikacjach naukowych).

## 2 Metody predykcji istotnych społecznie szeregów czasowych.

W rozdziale tym opisano opracowane metody do predykcji w szeregach czasowych liczby zakażeń i inflacji.

W pierwszej metodzie zastosowano multiregresję z wykorzystaniem pseudoinwersji Moore'a-Penrose'a, podając szczegółowy opis matematycznych fundamentów pseudoinwersji oraz wykorzystanych metryk do oceny precyzji prognoz. W kontekście implementacji, przedstawiono kluczowe metody w środowiskach programistycznych Python i MATLAB, co uwydatnia ich znaczenie dla praktycznej analizy danych. Podano również źródła danych i szczegółowo opisano proces ich przetwarzania, niezbędny do wykorzystania w badaniu. Zaprezentowanie matematycznego modelu metody, wsparte na przykładach macierzy i wektorów, umożliwiające zrozumienie jej działania. Dodatkowo rozszerzono model o zaproponowany wskaźnik gęstości

szczepień. Ograniczenia modelu zostały dokładnie omówione, co jest istotne dla zrozumienia jego zastosowań i potencjalnych ograniczeń.

Drugą metodą jest predykcja oparta na łańcuchach Markova. Metoda ta bazuje na danych jednego kraju. Przedstawiono model matematyczny klasycznego łańcucha Markova. Następnie w odniesieniu do danych o liczbie zakażeń przedstawiono sposób podziału na okna czasowe oraz na biny (przedziały wartości). Na podstawie segmentacji danych z danego okna czasowego do określonych binów przedstawiono utworzenie macierzy tranzycji oraz macierzy prawdopodobieństwa przejść. Opisano model predykcji z wykorzystaniem empirycznie określonych współczynników korygujących. Przedstawienie metody kończy podanie metryk określających dokładność predykcji z jej wykorzystaniem.

Trzecią metodą wykorzystaną do predykcji szeregów czasowych zakażeń jest metoda oparta na podobieństwie dynamiki zakażeń w  $n$ -najbliższych krajach-sąsiadach z wykorzystaniem średniej kroczącej. Metoda ta wymaga licznych predyktorów w postaci danych zakażeń z różnych krajów. Przedstawiono sposób przygotowania danych obejmujący normalizację wartości zakażeń. Kolejnym krokiem jest znalezienie określonej liczby krajów o najbardziej zbliżonej do prognozowanego kraju dynamice zakażeń. Zaprezentowano weryfikację danych wybranych krajów za pomocą metod statystycznych z uwzględnieniem metod graficznych. Następnie opisano model matematyczny oraz metryki pomiaru dokładności predykcji. Dodatkowo zaprezentowano metodę wizualizacji prognozy na zaproponowanym wykresie.

Kolejną grupą metod predykcji w szeregach czasowych są metody zastosowane do predykcji w szeregach czasowych inflacji. W grupie tej są dwie podstawowe metody (regresji wielomianem z optymalizacją parametrów i regresji wygładzania splajnem z optymalizacją parametrów) oraz ich rozszerzenia związane z optymalizacją. Trzecia metoda polega na korekcji błędu za pomocą regresji nieliniowej i jest stosowana w połączeniu z dwiema poprzednimi. Dla każdej z metod przedstawiono model matematyczny, algorytm z kluczowymi fragmentami implementacji w środowisku MATLAB oraz wizualizację działania na przykładowych danych. Wskazano metrykę do oceny dokładności predykcji na podstawie danych historycznych. W odróżnieniu od metod „pandemicznych”, w prognozowaniu inflacji wykorzystywane są dane jednego kraju.

## 2.1 Metoda multiregresji z wykorzystaniem pseudoinwersji Moore'a-Penrose'a.

Jedną z rozpatrywanych w tej pracy jest metoda oparta na modelu regresyjnym umożliwiającym predykcję na podstawie danych historycznych z uwzględnieniem błędu statystycznego. W metodzie zastosowano multiregresję w celu predykcji jednej zmiennej zależnej (w tym przypadku predykcji dobowych sumarycznych liczb zakażeń w danym kraju) na podstawie wybranych zmiennych niezależnych. W przypadku badania COVID-19 zmiennymi niezależnymi są szeregi czasowe dobowych zakażeń dla wybranych krajów na podstawie przyjętego kryterium uwzględniającego poziom szczepień i wielkość populacji w danym kraju.

Badaną zmienną zależną  $Y$  opisuje się wykorzystując omawianą tu metodę pseudoinwersji za pomocą jej modelu  $Y_m$  opisanego wzorem

$$Y_m = C_y \vec{A} \quad (14) \text{ opisanym w dalszej części paragrafu.}$$

Za kryterium oceny zastosowano APE i MAPE.

$$APE = \frac{|y - y_m|}{y} \times 100\% \quad (15)$$

MAPE jest uśrednioną wartością APE.

W kolejnych podrozdziałach przedstawiono cykl życiowy analizy danych prowadzący do predykcji zakażeń na COVID-19 z wykorzystaniem metody multiregresji. Cykl ten składa się z następujących etapów:

1. Akwizycja danych
2. Przygotowanie i preprocesowanie danych
3. Zaprojektowanie modelu
4. Utworzenie modelu
5. Odczyt uzyskanych wyników
6. Pomiar efektywności

Najistotniejszym równaniem w projektowanym modelu są przekształcenia macierzy niekwadratowych zawierających zmienne niezależne. Przekształcenia te wykonuje się za pomocą pseudoinwersji Moore'a-Penrose'a.

Pojęcie macierzy odwrotnej było niegdyś zastrzeżone wyłącznie dla macierzy kwadratowych. Uogólnione pojęcie macierzy odwrotnej w przypadku macierzy niekwadratowych wprowadzili niezależnie od siebie Eliakim Hastings Moore w 1920 r. i Roger Penrose w 1955 r.

Założmy, że równanie ma postać:

$$\vec{y} = A\vec{x} \quad (16)$$

Znając  $A$  oraz  $\vec{y}$  należy ustalić  $\vec{x}$ , gdzie  $\vec{y}$  i  $\vec{x}$  są wektorami, natomiast  $A$  jest macierzą.

W przypadku, kiedy  $A$  jest macierzą kwadratową, rozwiązanie wygląda następująco:

$$\vec{y} = A\vec{x} \quad (17)$$

$$A^{-1}\vec{y} = A^{-1}A\vec{x} \quad (18)$$

$$A^{-1}\vec{y} = I\vec{x} \quad (19)$$

$$A^{-1}\vec{y} = \vec{x} \quad (20)$$

W przypadku, kiedy macierz  $A$  nie jest macierzą kwadratową, nie można wykonać operacji  $A^{-1}$ . Jeśli spełniona jest nierówność  $m > n$ , gdzie  $m$  to liczba wierszy macierzy  $A$ ,  $n$  to liczba jej kolumn, wtedy można posłużyć się operacją  $A^+ \equiv$  pseudoinwersja.

Pseudoinwersja Moore'a-Penrose'a obliczana jest w następujący sposób:

$$A^+ = (A^T A)^{-1} A^T \quad (21)$$

Korzystając z właściwości  $A^+ A = I$  należy pamiętać, że  $AA^+ \neq I$  jeśli  $A$  nie podlega zwykłej inwersji.

Rozwiązanie równania w przypadku macierzy  $A$  nie-kwadratowej wygląda następująco:

$$\vec{y} = A\vec{x} \quad (22)$$

$$A^+\vec{y} \approx A^+A\vec{x} \quad (23)$$

$$A^+\vec{y} \approx I\vec{x} \quad (24)$$

$$A^+\vec{y} \approx \vec{x} \quad (25)$$

Należy zwrócić uwagę, że rozwiązanie równania za pomocą pseudoinwersji Moore'a-Penrose'a jest przybliżone i zawiera aproksymację rozwiązania dokładnego w sensie metody najmniejszych kwadratów.

W oprogramowaniu MATLAB wykonuje się powyższe przekształcenie poprzez zastosowanie operatora lewostronnego dzielenia macierzy zgodnie z zapisem:

$$A = X \backslash y'$$

W języku Python powyższe przekształcenie wykonuje się za pomocą metody biblioteki NumPy:

*linalg.pinv()*

Źródło: („numpy.linalg.pinv NumPy v1.23 Manual”, (2022)).

### 2.1.1 Przygotowanie i preprocesowanie danych.

Dane wykorzystane w metodzie pochodzą z „COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)”. W celu przeprowadzenia symulacji należy wstępnie przetworzyć pobrane dane. Przetworzenie danych polega na usunięciu zbędnych kolumn (tj. szerokość i długość geograficzna), pogrupowaniu danych z jednostki terytorialnej dla danego kraju (np. grupowanie po prowincji lub stanie), usunięciu danych regionalnych, sprowadzając do danych wyłącznie na poziomie krajowym.

Dane dotyczące kumulatywnej liczby zakażeń po pobraniu z repozytorium występują w formacie przedstawionym w Tab. 4.

Tab. 4. Format danych pobranych z CSSE przed wstępnym przetworzeniem. Źródło: opracowanie własne.

Province/State	Country/Region	Lat	Long	1/22/20	1/23/20
nan	Afghanistan	33.9391	67.71	0	0
nan	Albania	41.1533	20.1683	0	0
nan	Algeria	28.0339	1.6596	0	0
nan	Andorra	42.5063	1.5218	0	0
nan	Angola	-11.2027	17.8739	0	0
nan	Antarctica	-71.9499	23.347	0	0
nan	Antigua and Barbuda	17.0608	-61.7964	0	0
nan	Argentina	-38.4161	-63.6167	0	0
nan	Armenia	40.0691	45.0382	0	0
Australian Capital Territory	Australia	-35.4735	149.012	0	0
New South Wales	Australia	-33.8688	151.209	0	0
Northern Territory	Australia	-12.4634	130.846	0	0
Queensland	Australia	-27.4698	153.025	0	0
South Australia	Australia	-34.9285	138.601	0	0
Tasmania	Australia	-42.8821	147.327	0	0
Victoria	Australia	-37.8136	144.963	0	0
Western Australia	Australia	-31.9505	115.861	0	0
nan	Austria	47.5162	14.5501	0	0

Po wykonaniu operacji czyszczenia danych tabela z formatu określonego w Tab. 4. jest transponowana do macierzy  $Conf^T$  (Tab. 5. ) w formacie typowym dla prezentacji szeregów czasowych, w którym czas zmienia się w dół z kolejnymi wierszami, natomiast kolumny stanowią dane dla poszczególnych krajów.

Tab. 5. *Macierz Conf<sup>t</sup> przedstawiająca dane wybranych krajów pogrupowane w kolumnach. Źródło: opracowanie własne.*

Index	Afghanistan	Albania	Algeria	Andorra	Angola
2020-03-16 00:00:00	25	51	54	2	0
2020-03-17 00:00:00	26	55	60	39	0
2020-03-18 00:00:00	26	59	74	39	0
2020-03-19 00:00:00	26	64	87	53	0
2020-03-20 00:00:00	24	70	90	75	1
2020-03-21 00:00:00	24	76	139	88	2
2020-03-22 00:00:00	34	89	201	113	2
2020-03-23 00:00:00	40	104	230	133	3
2020-03-24 00:00:00	42	123	264	164	3
2020-03-25 00:00:00	74	146	302	188	3
2020-03-26 00:00:00	80	174	367	224	4
2020-03-27 00:00:00	91	186	409	267	4
2020-03-28 00:00:00	106	197	454	308	5
2020-03-29 00:00:00	114	212	511	334	7
2020-03-30 00:00:00	114	223	584	370	7

Kolejnym zbiorem danych wykorzystanym w tej metodzie są dane o szczepieniach pochodzące z „COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)”. Dane po wstępnym oczyszczeniu i przetworzeniu zapisane są w macierzy  $V_{ac}$ . W trakcie opracowania metody napotkano na problemy związane z danymi dotyczącymi liczby szczepień:

- zbiór danych o szczepieniach został stworzony później niż raportowano dane o zakażeniach
- dane dla poszczególnych krajów zaczęto raportować w różnych datach
- raportowano dane dla krajów sumarycznie oraz dla danych terytoriów odpowiadających podziałowi administracyjnemu danego kraju (np. dla prowincji, stanów, województw, itp.)

- dla wartości o indeksie  $beg$  i kraju  $C$  brakowało wartości liczbowych

### 2.1.2 Zaprojektowanie modelu

Państwa o wybranych numerach kolumn oznaczono jako  $C_1, C_2, \dots, C_k$  – to wektory pionowe. Wektory te tworzą okno czasowe  $w$  rozpoczynające się w  $i$ -tym dniu  $i = beg$  i kończą się w dniu  $i$ -tym  $i = fin$ . Rozmiar okna czasowego  $w$  określa się wzorem:

$$w = fin - beg \quad (26)$$

Wartości okna czasowego:

$$x_{11}, x_{21}, x_{31}, \dots, x_{w1} \in C_1; x_{12}, x_{22}, x_{32}, \dots, x_{w2} \in C_2; \dots; x_{1k}, x_{2k}, x_{3k}, \dots, x_{wk} \in C_k \quad (27)$$

Powstaje w ten sposób macierz  $C1$  (Tab. 6.) wraz z konkatelowanymi przesuniętym o  $hp$  wektorem kraju, dla którego będzie dokonywana predykcja.

Tab. 6. *Macierz C1 przedstawiająca dane wybranych krajów pogrupowane w kolumnach.*  
*Źródło: opracowanie własne.*

	Dania	Niemcy	USA	Polska
beg:	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	Y <sub>2</sub>
	X <sub>21</sub>	X <sub>22</sub>	X <sub>23</sub>	Y <sub>3</sub>
	X <sub>31</sub>	X <sub>32</sub>	X <sub>33</sub>	Y <sub>4</sub>
fin:	X <sub>41</sub>	X <sub>42</sub>	X <sub>43</sub>	Y <sub>5</sub>

Państwa wybiera się za pomocą kryterium określającego zbliżone wartości liczbowe na podstawie:

- Populacji
- Liczbie wykonanych szczepień

Kryterium, nazwane gęstość szczepień i oznaczone jako  $Vac_{Density}$  (ang. vaccination density – gęstość szczepień) mierzone w jednostce

kumulatywna liczba wykonanych szczepień  
tysiąc osobników populacji i obliczane dla każdego kraju na dzień odpowiadający

początkowi okna  $w$ , czyli wiersz o indeksie  $beg$  według wzoru:

$$Vac_{Density}(C) = Vac(beg, k) \div \frac{P}{1000} \quad (28) ,$$

gdzie:

$C$  – nazwa kraju, kolumna macierzy  $Vac$  o indeksie  $k$

$beg$  - numer wiersza macierzy  $Vac$  odpowiadający dacie początku okna  $w$

$Vac(beg, k)$  – wartość szeregu czasowego w dniu  $beg$  dla  $k$ -tej kolumny odpowiadającej liczbie szczepień wykonanych w  $C$  kraju

$P$  – populacja danego kraju

Gęstość szczepień dla wybranych krajów w dniu 1 stycznia 2021 r. została przedstawiona w Tab. 7. Krajem o największej gęstości szczepień zajmującym pierwsze miejsce w tabeli jest Izrael, którego gęstość szczepień wynosi  $\sim 129,45$ . Izrael na przełomie 2020 i 2021 r. był liderem w liczbie zaszczepionych mieszkańców (Rossman et al. (2021)). Krajem o najmniejszej niezerowej gęstości szczepień jest ostatnia w tabeli Francja. Wartość gęstości szczepień dla Francji w rozpatrywanym dniu wynosi 0,0069.

Jeśli predykcja z uwzględnieniem rozpatrywanej metody będzie dotyczyć Polski i okno w rozpoczyna się 1 stycznia 2021, zgodnie i założeniami: krajami – predyktorami będą Estonia, Włochy, Oman, Łotwa, Czechy i Argentyna.

W związku z napotkanymi problemami w zbiorze danych dotyczących szczepień, wprowadzono ograniczenia i przyjęto założenia:

1. Podobieństwo liczby wykonanych szczepień badamy na poziomie kraju, dla którego takie dane są zawarte w macierzy  $Vac$  w dniu o indeksie  $beg$ . W macierzy  $Vac$  pozostawiono tylko kolumny odpowiadające wartościom dla danych krajów. W przypadku braku wartości liczbowej w dniu o indeksie  $beg$  usunięto taki kraj z macierzy  $Vac$ .
2. Uwzględnia się po 3 kraje o większej oraz mniejszej gęstości szczepień.
3. Jeśli nie ma 3 krajów o mniejszej gęstości szczepień (dla 3 krajów o najmniejszej gęstości szczepień), to uwzględniamy tylko 3 kraje o większej gęstości szczepień.

4. Jeśli nie ma 3 krajów o większej gęstości szczepień, to uwzględniamy mniejszą liczbę dostępnych krajów o większej gęstości szczepień (przypadek dla 3 krajów o największej gęstości szczepień)

Ograniczeniami metody są:

1. zakres dat obejmujący niezerowe wartości liczbowe dotyczące szczepień w co najmniej 5 krajach łącznie z krajem, dla którego będziemy dokonywali predykcji.
2. Dostępność danych w zakresie dat w liczbie odpowiadającej co najmniej sumie  $w+hp$ .

Tab. 7. Tabela przedstawiające wielkość populacji, liczbę wykonanych szczepień w dniu odpowiadającym indeksowi *beg* macierzy *Vac* (w tym przypadku 1 styczeń 2021 r.) oraz gęstość szczepień (kolumna *Vac\_Density*). Źródło: opracowanie własne.

Kraj	Populacja	beg = 1.01.2021	Vac Density
Israel	8655541	1120504	129,4551
Bahrain	1701583	60097	35,3183
United Kingdom	67886004	1001985	14,7598
Iceland	341250	4875	14,2857
US	329466283	2786421	8,4574
Denmark	5792203	47246	8,1568
Russia	145934460	800000	5,4819
Slovenia	2078932	11224	5,3989
Lithuania	2722291	9747	3,5804
Portugal	10196707	32749	3,2117
China	1404676330	4500000	3,2036
Germany	83783945	255112	3,0449
Canada	37855702	103790	2,7417
Luxembourg	625976	1200	1,9170
Croatia	4105268	7864	1,9156
Estonia	1326539	2487	1,8748
Italy	60461828	88171	1,4583
Oman	5106622	7231	1,4160
Poland	37846605	50391	1,3315
Latvia	1886202	2259	1,1976
Czechia	10708982	12157	1,1352
Argentina	45195777	32013	0,7083
Romania	19237682	13242	0,6883
Bulgaria	6948445	4739	0,6820
Kuwait	4270563	2500	0,5854
Austria	9006400	5017	0,5570
Hungary	9660350	5110	0,5290
Chile	19116209	8649	0,4524
Norway	5421242	2183	0,4027
Ireland	4937796	1800	0,3645
Finland	5540718	1767	0,3189
Greece	10423056	3149	0,3021
Malta	441539	130	0,2944
Sweden	10099270	2253	0,2231
Mexico	127792286	24998	0,1956
Belgium	11589616	794	0,0685
Costa Rica	5094114	55	0,0108
France	65273512	449	0,0069

Z macierzy *C1* (Tab. 6.) tworzy się macierz *C* (Tab. 8. i Tab. 9.) o wymiarach [ $w \times k$ ], gdzie *k* to liczba wybranych krajów.

Tab. 8. Macierz  $C$  ze zmienną objaśnianą jako wektor  $Y$  w notacji Ivakhnenko (Ivakhnenko et al. 1976) ze wskazaniem horyzontu prognozy. Źródło: opracowanie własne.

$C_1$	$C_2$	$C_3$	$Y$
•	•	•	
•	•	•	•
•	•	•	•
•	•	•	•
			•

↑hp  
↓hp

Tab. 9. Macierz  $C$  zawierająca predyktory – wektory skumulowanej liczby zakazeń w krajach  $C_k$  w oknie czasowym w pomiędzy datami beg i fin. Źródło: opracowanie własne.

	$C_1$	$C_2$	...	$C_K$
beg:	$X_{1,1}$	$X_{1,2}$	...	$X_{1,k}$
	$X_{2,1}$	$X_{2,2}$	...	$X_{2,k}$
	...	...	...	...
fin:	$X_{w,1}$	$X_{w,2}$	...	$X_{w,k}$

Dodatkowo tworzy się wektor  $\vec{y}$  (0) reprezentujący państwo będące zmienną objaśnianą  $Y$  (0). Wektor  $y$  należy przesunąć w przód o horyzont prognozy  $hp$  w celu umożliwienia predykcji.

Tab. 10. Wektor  $\vec{y}$  reprezentujący państwo będące zmienną objaśnianą  $Y$ , który należy przesunąć o horyzont prognozy  $hp$  w celu zbudowania modelu i umożliwienia predykcji.

	$\vec{y}$
beg+hp:	$Y_{hp}$
	$Y_{hp+1}$
	...
fin+hp:	$Y_{hp+w}$

Wykonanie pseudoinwersji Moore'a-Penrose'a wymaga odpowiednich wymiarów macierzy (Katsikis i Pappas (2008), Ben-Israel (2002)). Wobec tego, do macierzy  $C$  konkatelowane są wektory jedynek  $\mathbf{1}$  (Tab. 11. ).

Tab. 11. Macierz  $C$  oraz konkatelowany wektor  $\mathbf{1}$ . Źródło: opracowanie własne.

$$\begin{array}{l} \text{beg:} \\ \text{fin:} \end{array} \left[ \begin{array}{cccc} \mathbf{1} & C_1 & C_2 & \dots & C_k \\ 1 & X_{1,1} & X_{1,2} & \dots & X_{1,k} \\ 1 & X_{2,1} & X_{2,2} & \dots & X_{2,k} \\ 1 & \dots & \dots & \dots & \dots \\ 1 & X_{w,1} & X_{w,2} & \dots & X_{w,k} \end{array} \right]$$

Macierz  $C$  przyjmie postać  $C = [\mathbf{1}, C_1, C_2, \dots, C_k]$

gdzie  $C_1, C_2, \dots, C_k$  to wektory kolumnowe zawierające dane wybranych krajów.

Równanie opisujące model przyjmuje postać:

$$\vec{y} = C\vec{A} \quad (29)$$

Znając wektor  $\vec{y}$  oraz dysponując macierzą  $C$  (Tab. 12. ), brakującym elementem równania jest wektor  $\vec{A}$ . Wektor ten można obliczyć wykorzystując pseudoinwersję Moore'a-Penrose'a macierzy  $C$  (której wynikiem jest macierz oznaczona  $C^+$ ) i sprowadzając równanie do postaci:

$$C^+\vec{y} \approx \vec{A} \quad (30)$$

Tab. 12. Zobrazowanie szeregów czasowych w modelu. Macierz  $C$  oznaczono kolorem pomarańczowym, natomiast wektor  $\vec{y}$  przesunięty o  $hp=7$  kolorem żółtym. Źródło: opracowanie własne.

		Ones	Bahrain	United Kingdom	Iceland	US	Russia	Slovenia	Lithuania	Y
beg	1.01.2021	1	92913	2549671	5754	20398843	3153960	123950	147862	166567
	2.01.2021	1	93184	2607548	5754	20671281	3179898	124343	149105	168182
	3.01.2021	1	93478	2662664	5754	20874515	3203743	125086	150529	169358
	4.01.2021	1	93766	2721583	5795	21061497	3226758	125858	151719	171436
	5.01.2021	1	93995	2782676	5811	21293669	3250713	128370	153763	173433
	6.01.2021	1	94284	2845230	5832	21553509	3274615	131724	156558	175663
	7.01.2021	1	94633	2898028	5845	21838560	3297833	134545	158856	177507
	8.01.2021	1	95030	2966224	5867	22141563	3321163	136629	160827	179170
	9.01.2021	1	95317	3026299	5880	22405007	3344175	138518	162531	180915
	10.01.2021	1	95558	3081319	5890	22614839	3366715	139281	163989	182161
	11.01.2021	1	95879	3127617	5898	22819595	3389733	139713	164621	183400
	12.01.2021	1	96195	3173273	5912	23041963	3412390	141587	165828	184476
	13.01.2021	1	96470	3220936	5935	23270690	3434934	143679	167607	185835
fin	14.01.2021	1	96812	3269737	5948	23505723	3459237	145472	168893	185836
	15.01.2021	1	97020	3325623	5956	23749964	3483531	147016	170079	187998
	16.01.2021	1	97268	3367060	5956	23966962	3507201	148556	171225	188878
	17.01.2021	1	97607	3405734	5956	24143110	3530379	149125	172002	189767
	18.01.2021	1	97940	3443344	5970	24284465	3552888	149424	172452	190574
	19.01.2021	1	98260	3476787	5970	24441483	3574330	151137	173282	191299
	20.01.2021	1	98573	3515776	5975	24629169	3595136	152851	174497	192187
	21.01.2021	1	98878	3553757	5981	24821335	3616680	154306	175745	192947

Obliczony wektor  $\vec{A}$  współczynników równania regresyjnego pozwala na wyznaczenie modelu wektora  $\vec{y}$  oznaczonego jako  $Y_m$ . Należy zwrócić uwagę na przesunięcie  $Y_m$  o horyzont predykcji  $hp$  zobrazowany w Tab. 8. Równanie regresyjne przyjmuje postać:

$$Y_m = C_y \vec{A} \quad (31)$$

gdzie:

$Y_m$  to model zmiennej objaśnianej,

$C_y$  – to macierz powstała w analogiczny sposób jak macierz  $C$  z tym, że indeks wierszy szeregu czasowego jest przesunięty w czasie w przód o liczbę wierszy odpowiadający wartości horyzontu predykcji  $hp$ .

Zobrazowanie przykładowych danych  $Y_m$  i  $C_y$  równania regresyjnego przedstawiono w Tab.

13.

Tab. 13. Zobrazowanie szeregów czasowych w modelu na etapie predykcji wektora  $Y_m$  o horyzoncie predykcji  $hp=7$  począwszy od daty 22.01.2021 (oznaczony kolorem zielonym). Macierz  $C_y$  oznaczono kolorem pomarańczowym. Źródło: opracowanie własne.

		Ones	Bahrain	United Kingdom	Iceland	US	Russia	Slovenia	Lithuania	Ym
beg	1.01.2021	1	92913	2549671	5754	20398843	3153960	123950	147862	
	2.01.2021	1	93184	2607548	5754	20671281	3179898	124343	149105	
	3.01.2021	1	93478	2662664	5754	20874515	3203743	125086	150529	
	4.01.2021	1	93766	2721583	5795	21061497	3226758	125858	151719	
	5.01.2021	1	93995	2782676	5811	21293669	3250713	128370	153763	
	6.01.2021	1	94284	2845230	5832	21553509	3274615	131724	156558	
	7.01.2021	1	94633	2898028	5845	21838560	3297833	134545	158856	
	8.01.2021	1	95030	2966224	5867	22141563	3321163	136629	160827	
	9.01.2021	1	95317	3026299	5880	22405007	3344175	138518	162531	
	10.01.2021	1	95558	3081319	5890	22614839	3366715	139281	163989	
	11.01.2021	1	95879	3127617	5898	22819595	3389733	139713	164621	
	12.01.2021	1	96195	3173273	5912	23041963	3412390	141587	165828	
	13.01.2021	1	96470	3220936	5935	23270690	3434934	143679	167607	
fin	14.01.2021	1	96812	3269737	5948	23505723	3459237	145472	168893	
hp	15.01.2021	1	97020	3325623	5956	23749964	3483531	147016	170079	Ym[1]
	16.01.2021	1	97268	3367060	5956	23966962	3507201	148556	171225	Ym[2]
	17.01.2021	1	97607	3405734	5956	24143110	3530379	149125	172002	Ym[3]
	18.01.2021	1	97940	3443344	5970	24284465	3552888	149424	172452	Ym[4]
	19.01.2021	1	98260	3476787	5970	24441483	3574330	151137	173282	Ym[5]
	20.01.2021	1	98573	3515776	5975	24629169	3595136	152851	174497	Ym[6]
	21.01.2021	1	98878	3553757	5981	24821335	3616680	154306	175745	Ym[7]
hp	22.01.2021									Ym[8]
	23.01.2021									Ym[9]
	24.01.2021									Ym[10]
	25.01.2021									Ym[11]
	26.01.2021									Ym[12]
	27.01.2021									Ym[13]
	28.01.2021									Ym[14]

### 2.1.3 Odczyt uzyskanych wyników

Wartość predykcijną zmiennej  $y$  w następnych krokach szeregu czasowego w zakresie  $fin+hp$ :  $fin + 2 \times hp$  odczytujemy w wyniku obliczenia równania regresyjnego w elementach o indeksach  $[hp..2 \times hp]$  w wektorze  $Y_m$  (oznaczone kolorem zielonym w Tab. 13. ).

## 2.1.4 Pomiar efektywności

Symulacje przeprowadzone dla różnych warunków (np. w zależności od wartości parametrów w lub hp) dla i-tego dnia porównuje się za pomocą APE:

$$APE(i) = \frac{|Y(i) - Y_m(i)|}{Y(i)} \times 100\% \quad (32),$$

gdzie:

$Y$  – wartości rzeczywiste dla i-tego dnia

$Y_m$  – wartość predykcji dla i-tego dnia

Efektywność symulacji przeprowadzonych dla tych samych parametrów, ale w różnych datach mierzy się za pomocą MAPE:

$$MAPE = \frac{1}{i} \sum_1^i APE(i) \quad (33)$$

gdzie  $i$  – liczba symulacji.

W przypadku opisywanej metody liczba symulacji równa jest liczbie dni w horyzoncie predykcji:  $i = hp$

## 2.2 Metoda oparta na łańcuchach Markowa.

### 2.2.1 Wprowadzenie

Kolejną z rozpatrywanych w tej pracy metod predykcji zakażeń COVID-19 jest metoda oparta na łańcuchach Markowa. Model Markowa wywodzi się od rosyjskiego matematyka Andrieja Andriejewicza Markowa (1856 – 1922), który zajmował się teorią liczb oraz analizy i teorii prawdopodobieństwa. Markow rozszerzył prawo wielkich liczb i centralne twierdzenie graniczne o pewnych sekwencjach, zależnych zmiennych losowych tworzących specjalne klasy współcześnie znanych łańcuchów Markowa.

Łańcuchem Markowa o zbiorze stanów  $S \subseteq \mathbb{R}$  nazywamy ciąg zmiennych losowych  $X_0, X_1, X_2, \dots$  taki, że

$$P(X_n = x_n | X_{n-1} = x_{n-1} \wedge X_{n-2} = x_{n-2} \wedge \dots \wedge X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1}) = p_{x_{n-1}, x_n} \quad (34)$$

dla każdego  $n > 0$  i ciągu stanów  $x_0, x_1, \dots, x_n \in S$ .

W tej definicji zmienna  $X_t$  opisuje stan, w którym znajduje się łańcuch Markowa w chwili  $t$ . Warunek z definicji łańcucha Markowa mówi, że zmienna  $X_t$  zależy tylko od zmiennej  $X_{t-1}$ . W zależności od  $n$  zależności od poprzednich stanów mówimy o modelu Markowa  $n$ -stopnia.

Baum i Petrie (1966) opisali rozwinięcie metody Markowa pod pojęciem ukrytego modelu Markowa (HMM). Model ten jest procesem stochastycznym, który nie jest obserwowalny, bowiem jest ukryty. Można jednak go zaobserwować tylko poprzez inny zestaw procesów stochastycznych, które wytwarzają ciąg obserwacji. Natomiast obserwacja jest probabilistyczną funkcją stanu.

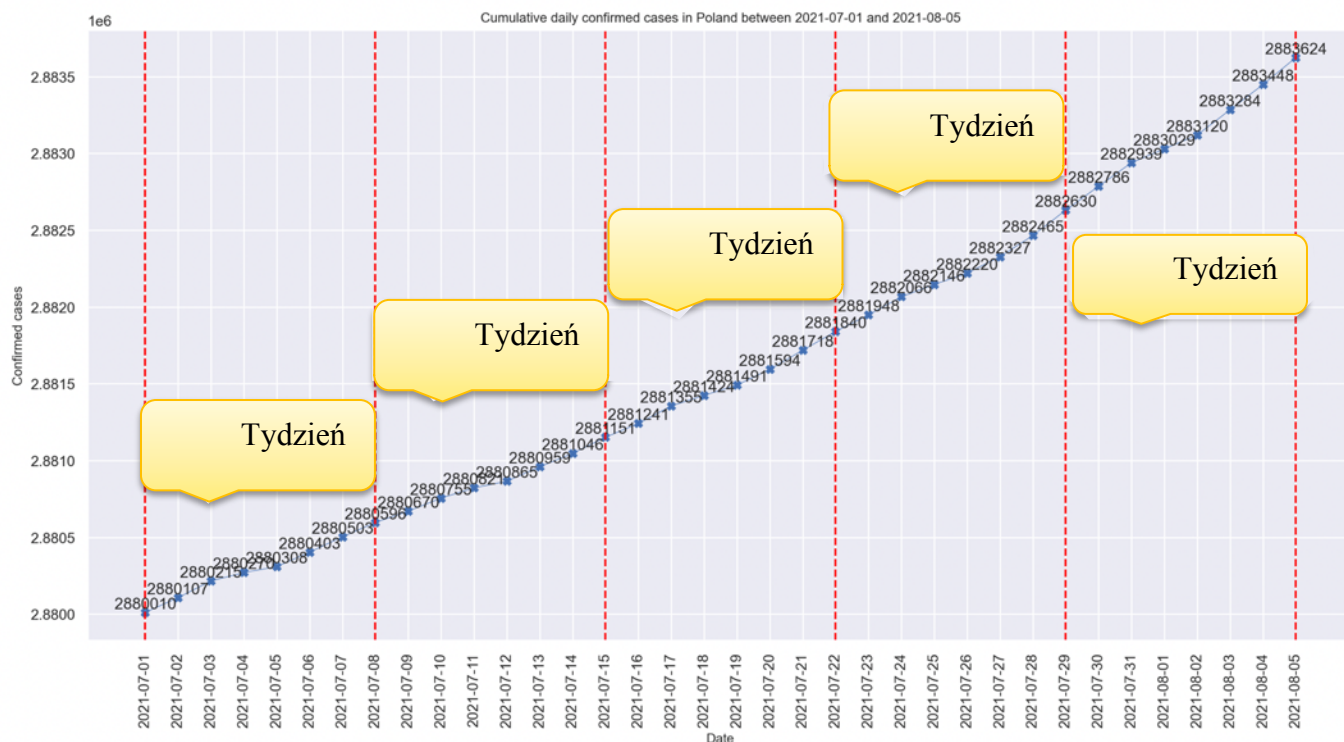
Na potrzeby przeprowadzenia predykcji liczby zakażeń COVID-19 skorzystano z własności stosowanej dla dyskretnych procesów Markowa:

$$P(X_{n+1} \leq y | X_0, X_1, X_2 \dots X_n) = P(X_{n+1} \leq y | X_n) \quad (35)$$

Predykcja rozprzestrzeniania się COVID-19 w opisywanej metodzie dzieli się na etapy:

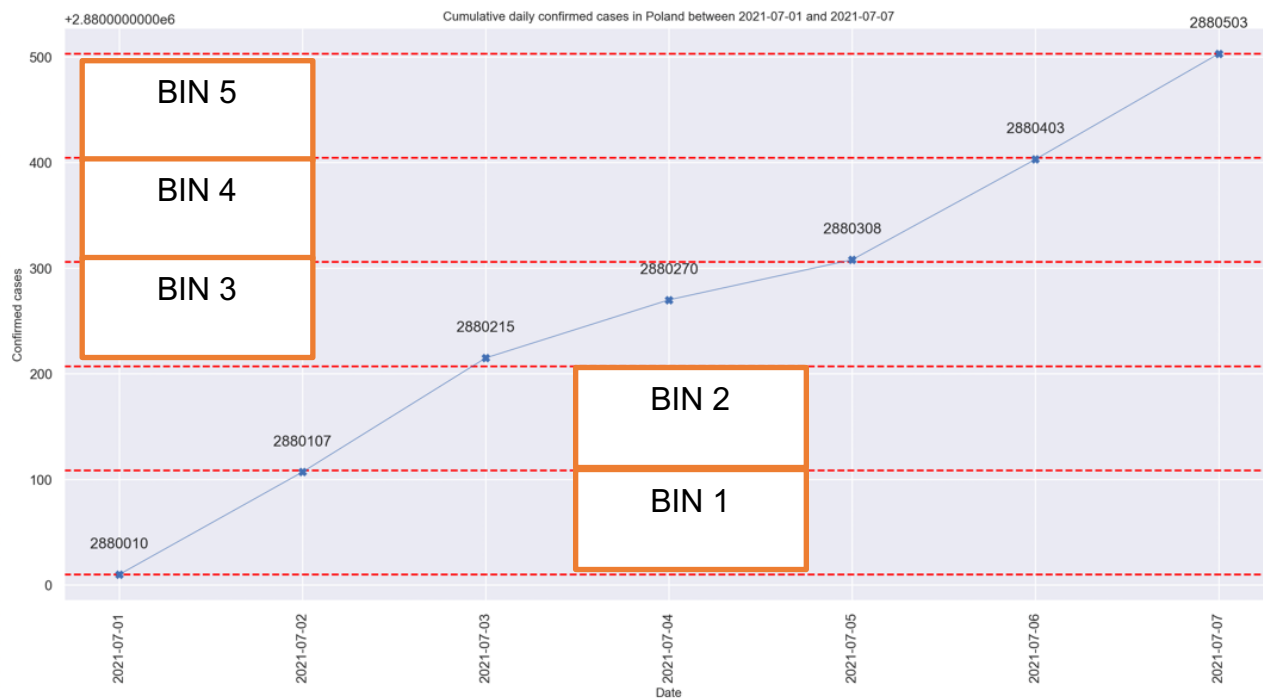
1. **Rozkład początkowy.** Zmienna  $X_0$  w postaci np. liczby zakażeń rozkładana jest na dyskretne przedziały zależne od czasu.
2. Dokonanie klasyfikacji stanów.
3. **Stworzenie macierzy przejść.** Przyjęto zasadę podziału wszystkich zakażeń wg kryteriów wielkości szeregu czasowego i liczby okien obserwacji. Następnie dla każdego szeregu czasowego dokonano kwantyzacji na 5 równych stref nazwanych 1/2/3/4/5 bin , gdzie 1 bin to minimalna wartość, a 5 bin to wartość maksymalna.
4. **Stworzenie macierzy tranzycji** w celu określenia prawdopodobieństwa przejścia do poszczególnych binów z poprzednich binów.
5. **Predykcja liczby zakażeń** na podstawie macierzy tranzycji.
6. **Określenie błędu APE** oraz MAPE dla każdego punktu. Sytuacją pożądaną dla celów dokładności predykcji jest tendencja malejąca wartości MAPE.

W celu dokonania rozkładu początkowego należy wybrać badany okres, wskazując daty początkową i końcową. Ze względu na okresowe zakłócenia danych spowodowane dniami wolnymi od pracy (weekendy), zaleca się, aby zakres dat odzwierciedlał okres odpowiadający wielokrotności 7 dni (Rys. 18. ).



Rys. 18. Podział wybranego zakresu dat na 7 dniowe okna czasowe. Źródło: opracowania własne.

Okna czasowe badanego szeregu czasowego dzieli się na 5 binów (Rys. 19. ). Są to równe przedziały wartości pomiędzy wartością minimalną (wartość minimalna bina 1) a wartością maksymalną (wartość maksymalna bina 5) szeregu czasowego w rozpatrywanym oknie czasowym.



Rys. 19. Podział okna czasowego na 5 binów. Źródło: opracowania własne.

Istotne jest zróżnicowanie danych w każdym oknie czasowym. W przypadku różnicy pomiędzy maksymalną a minimalną wartością mniejszą niż liczba binów nie jest możliwy podział na biny. Również należy brać pod uwagę zaokrąglenia przy dzieleniu bez reszty w samym algorytmie, przy klasyfikacji danego elementu do bina. Przyjęto empirycznie, że minimalne zróżnicowanie w oknie czasowym musi wynosić co najmniej 20.

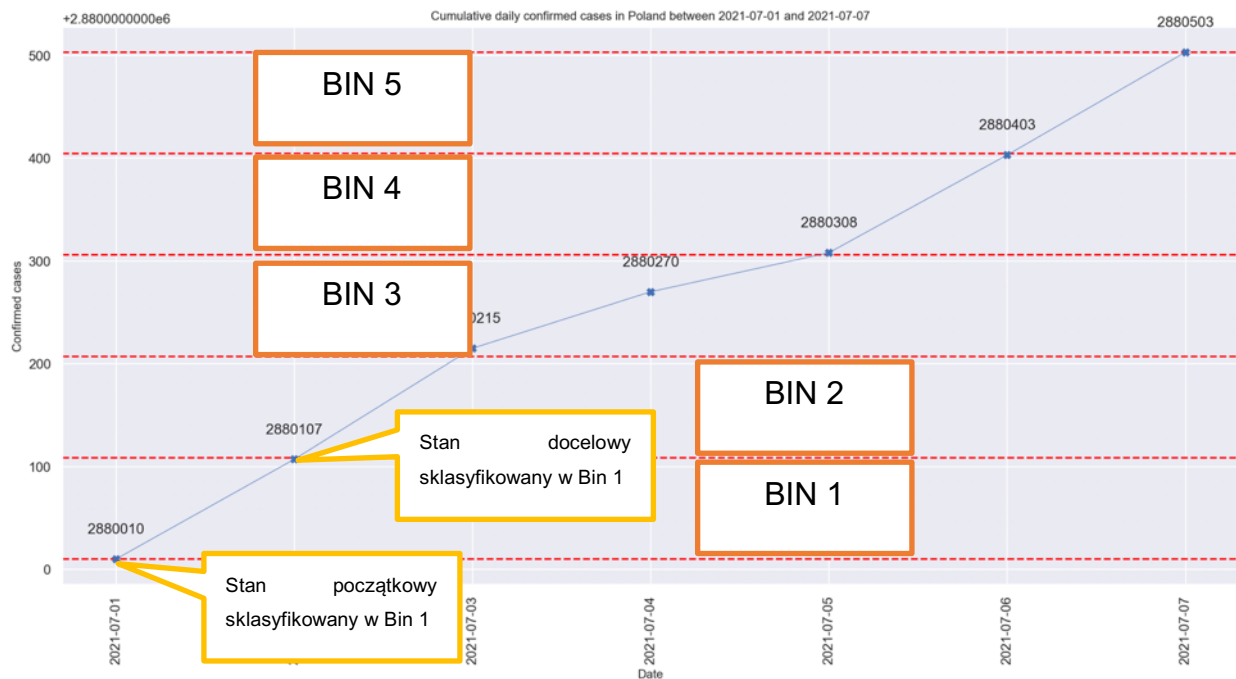
Problemem jest ostatni dzień okna czasowego, ponieważ okno czasowe się kończy i nie można badać przejścia – brak jest stanu docelowego. Rozwiązaniem tego problemu jest przesunięcie kolejnego okna czasowego o 1 dzień wstecz (Rys. 20. ). Czyli ostatni dzień okna czasowego jest pierwszym dniem kolejnego okna czasowego. W ten sposób nie traci się cennej informacji o przejściu z tego dnia do kolejnego bina. Wiersz zawierający ostatni dzień okna czasowego oraz brak informacji o przejściu jest usuwany ze zbioru danych i zastępowany informacją z kolejnego okna czasowego

Index	Poland	bin no	next bin	tw number	bin mid
2021-08-28	2888028	1	1	1	
2021-08-29	2888231	1	2	1	
2021-08-30	2888385	2	2	1	174
2021-08-31	2888539	3	3	1	174
2021-09-01	2888693	4	4	1	174
2021-09-02	2889412	4	5	1	174
2021-09-03	2889773	5	nan	1	174
2021-09-03	2889773	1	1	2	234
2021-09-04	2890161	1	2	2	
2021-09-05	2890549	2	2	2	
2021-09-06	2890937	2	3	2	234
2021-09-07	2891071	3	4	2	234
2021-09-08	2891602	4	5	2	234
2021-09-09	2892113	5	nan	2	234

Rys. 20. *Zobrazowanie przesunięcia kolejnego okna czasowego (tw=2) w celu uzyskania informacji o przejściu pomiędzy binami (tutaj informacja z dnia 03.09.2021 r.). Źródło: opracowania własne.*

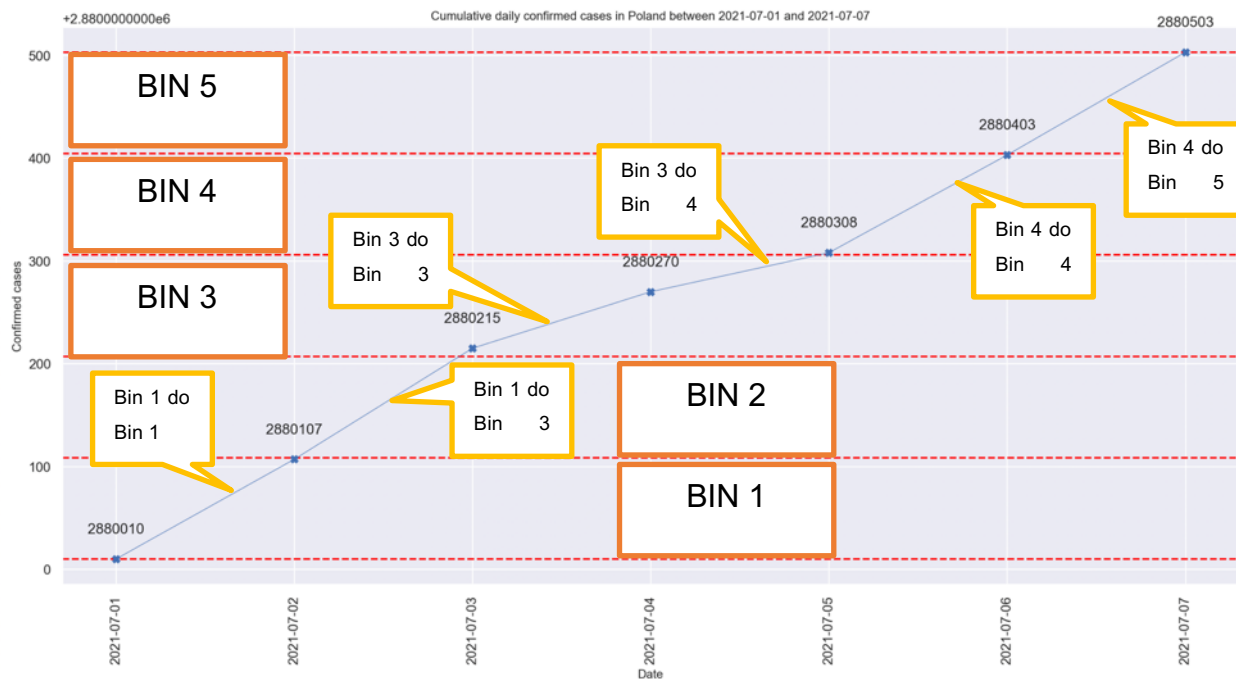
Liczba binów została określona eksperymentalnie i jest optymalna ze względu na dokładność predykcji oraz rozdzielczość danych wejściowych. Próby wykorzystania 3 binów skutkowały niezadowalającą precyzją informacji o liczbie zakażeń w danym binie w stosunku do całego przedziału. Większa liczba binów niż 5, np. 7 powoduje obniżenie dynamiki przejścia stanów i utworzenie macierzy tranzycji w formie nie nadającej się do dokonania predykcji zakażeń. Innymi słowy w poszczególnym binie zaistniałoby zbyt mało zdarzeń, aby określić prawdopodobieństwo wystąpienia tego stanu.

Macierz przejść  $M$  zawiera informację ilościową dotyczącą zmian lub braku zmian binów dla kolejnych elementów szeregu czasowego (Rys. 21. ). W danym przejściu uwzględnia się wyłącznie 2 wyrazy szeregu czasowego: początkowy i końcowy. Macierz przejść prowadzi do utworzenia macierzy tranzycji.



Rys. 21. Zobrazowany przykład określania przejścia dla 2 kolejnych stanów (początkowego i końcowego). Źródło: opracowanie własne.

Po dokonaniu klasyfikacji przejścia (np. z Bin 1 do Bin 1 zgodnie z Rys. 21. ) klasyfikuje się przejście w macierzy  $M$  poprzez inkrementację wartości w elemencie macierzy  $M_{ij}$ , gdzie  $i$  to numer kolumny odpowiadający numerowi bina źródłowego, natomiast  $j$  to numer wiersza odpowiadający numerowi bina docelowego. Po dokonaniu klasyfikacji i określenia przejścia (Rys. 22. i Rys. 23. ) rozpatruje się kolejne elementy szeregu czasowego, czyli poprzednio rozpatrywany stan docelowy jest rozpatrywany jako stan początkowy, a następujący po nim element szeregu czasowego jest elementem docelowym.



Rys. 22. Zobrazowana klasyfikacja stanów i określenie przejść w oknie czasowym.  
 Źródło: opracowanie własne.

Ostatecznie macierz przybiera postać przedstawioną na Rys. 23.

		Bin źródłowy						
		Bin 1	Bin 2	Bin 3	Bin 4	Bin 5		
M=	83	119	1	0	0	Bin 1	Bin docelowy	
	0	69	111	8	0	Bin 2		
	0	0	36	106	6	Bin 3		
	0	0	0	10	114	Bin 4		
	0	0	0	0	57	Bin 5		

Rys. 23. Przykładowa macierz przejść dla danego szeregu czasowego wraz z objaśnieniem stosowanej notacji. Źródło: opracowania własne.

Macierz tranzycji zwana macierzą prawdopodobieństw P wyznacza się wg wzoru:

$$P_{ij} = \frac{M_{ij}}{\sum_{j=1}^5 M_{ij}} \quad (36),$$

gdzie  $i$  to numer wiersza, a  $j$  to numer kolumny.

Macierz ta przedstawia prawdopodobieństwo przejścia między binami (Rys. 24. ). Macierz prawdopodobieństwa ma formę macierzy Markowa - stochastycznie prawej. Oznacza to, że suma

elementów każdego wiersza równa jest 1 oraz jej wszystkie elementy przyjmują wartości nieujemne.

P =

0,4089	0,5862	0,0049	0	0
0	0,3670	0,5904	0,0426	0
0	0	0,2432	0,7162	0,0405
0	0	0	0,0806	0,9194
0	0	0	0	1

Rys. 24. Zobrazowana macierz prawdopodobieństw  $P$  w formie mapy ciepła (zielone odcienie przyjmują wartości wyższe, żółte wartości bliskie medianie oraz czerwone minimalne). Pogrubioną czcionką oznaczono wartości maksymalne w danych wierszach. Źródło: opracowania własne.

Następnym krokiem jest wybranie elementów o maksymalnych wartościach prawdopodobieństwa w każdym wierszu. Na tej podstawie będzie podejmowana decyzja o przejściu do konkretnego bina w kolejnym kroku. Wiersze odpowiadają numerom binów źródłowych, natomiast kolumny numerom binów docelowych. W przypadku identycznych wartości dla danego wiersza wybiera się ostatnią kolumnę. Numerem bina docelowego jest numer kolumny zawierającej element wiersza (o numerze odpowiadającym numerowi bina źródłowego) o największej wartości prawdopodobieństwa (Rys. 25. ).

P =

0,4089	0,5862	0,0049	0	0
0	0,3670	0,5904	0,0426	0
0	0	0,2432	0,7162	0,0405
0	0	0	0,0806	0,9194
0	0	0	0	1

Z Bin 1 do Bin 2

Z Bin 2 do Bin 3

Z Bin 3 do Bin 4

Z Bin 4 do Bin 5

Z Bin 5 do Bin 5

Rys. 25. Zobrazowanie przejść pomiędzy binami na podstawie wartości maksymalnych prawdopodobieństwa w macierzy prawdopodobieństw. Pogrubioną czcionką oznaczono wartości maksymalne w danych wierszach. Źródło: opracowania własne.

### 2.2.2 Predykcja liczby zakażeń.

Predykcja zakażeń odbywa się na podstawie prawdopodobieństwa przejścia do kolejnego bina określonym w macierzy P oraz przyroście dziennych zakażeń. Przyrost dziennych zakażeń  $c_m$  określa się wg wzoru:

$$c_m = 0,52 \times (b_d - b_z) \times C_b + 0,92 \times C_b \quad (37)$$

gdzie:

$b_z$  – numer bina źródłowego

$b_d$  – numer prognozowanego bina docelowego

$C_b$  – połowa liczby zakażeń w jednym binie w danym oknie czasowym

Współczynniki 0,52 oraz 0,92 zostały dobrane empirycznie w wyniku przeprowadzonych symulacji.

### 2.2.3 Określenie błędu APE i MAPE.

Dokładność działania metody zostanie oceniona pod względem APE i MAPE. Im mniejsze wartości, tym model jest bardziej precyzyjny.

$$APE(i) = \left| \frac{c_r(i) - c_m(i)}{c_r(i)} \right| \times 100\% \quad (37-1),$$

gdzie:

$c_r$  – liczba rzeczywistych zakażeń w i-tym dniu

$c_m$  – liczba zakażeń opracowana wg modelu Markowa

W przypadku  $c_r(i) = 0$ , czyli kiedy brak jest dziennego przyrostu w i-tym dniu, APE nie może być obliczone, bowiem zachodzi iloraz przez 0. W tym przypadku APE(i) jest pomijane w dalszej analizie.

Znając wartości APE, można wyliczyć MAPE:

$$MAPE = \frac{1}{i} \sum_1^i APE(i) \quad (38),$$

gdzie:

$i$  - liczba symulacji, dla  $APE(i) \neq 0$

## 2.3 Metoda oparta na podobieństwie dynamiki zakażeń w n-najbliższych sąsiadach z wykorzystaniem średniej kroczącej.

Prezentowana metoda umożliwia zastosowanie predykcji dziennych zakażeń dla danego kraju bazując na dobowych przyrostach zakażeń na podstawie podobieństw z danymi innych krajów. W metodzie tej istotne jest dobranie właściwych danych oraz określenie dynamiki rozwoju pandemii w danym kraju z podziałem na fazy.

Długość okna uczenia powinna wynosić minimum dwukrotność długości horyzontu predykcji. W wyjaśnieniu metody przyjęto horyzont predykcji 12 dni i okno uczenia w poprzedzających 24 dniach.

### 2.3.1 Przygotowanie danych liczbowych

Danymi wejściowymi do metody są dane zawierające szeregi czasowe skumulowanych wartości dobowych zakażeń zapisane w macierzy *Conf* (od ang. Confirmed – potwierdzone). Zaznaczyć należy, że wartości dobowych skumulowanych przypadków zakażeń nie mają wartości ujemnych, co zostanie dalej uwzględnione w metodzie.

Przygotowanie danych do zastosowania w metodzie sprowadza się do sekwencji czynności:

1. Pobranie dostępnego on-line zbioru danych z CSSE.
2. Usunięcie zbędnych i redundantnych kolumn ze zbioru.
3. Transpozycja zbioru do postaci macierzy *Conf*.
4. Stworzenie podzbioru macierzy *Conf* zawierającej segment danych reprezentujący 24 dni poprzedzające horyzont predykcji.
5. Normalizacja danych.
6. Wydzielenie z macierzy *Conf* wektora zawierającego dane dla kraju bazowego.

#### **Pobranie danych on-line.**

Dane pobierane są w formacie pliku CSV za pomocą stworzonego skryptu w języku Python z GitHub CSSE z lokalizacji <https://raw.githubusercontent.com/CSSEGISandData/COVID->

19/master/csse\_covid\_19\_data/csse\_covid\_19\_time\_series/time\_series\_covid19\_confirmed\_global.csv .

### **Usunięcie zbędnych i redundantnych kolumn ze zbioru.**

Pobrane dane zawierają informację redundantną lub zbędną dla analizy. Taką informacją jest przebieg zakażeń występujący lokalnie i w określonym krótkim terminie. Przykładem jest przebieg zakażeń na wycieczkowcu „Diamond Princess” (Jimi, H., & Hashimoto, G. (2020)) lub na olimpiadzie zimowej w Tokyo (Huang et al.(2020)). Ponadto ze zbioru danych usuwane są dane regionalne, które statystycznie stanowią cenną informację, ale i tak są pogrupowane w danych krajowych.

### **Transpozycja zbioru i utworzenie segmentu danych.**

Z macierzy *Conf* wybiera się podzbiór danych tworzących segment danych z 24 dni poprzedzający horyzont predykcji (Tab. 14. ).

Tab. 14. Przygotowanie danych wejściowych do algorytmu Macierz Conf zawierająca 24 dniowy segment danych (skumulowane dobowe przypadki zakażeń) poprzedzający horyzont predykcji. Źródło: opracowania własne.

Index	Afghanistan	Albania	Algeria	Andorra	Angola	Antarctica	igua and Barbi	Argentina	Armenia	Australia	Austria
2022-03-27	177520	273432	265629	39713	99115	11	7485	9026075	422458	4326181	3734990
2022-03-28	177602	273432	265641	39713	99115	11	7491	9028730	422468	4381919	3760649
2022-03-29	177658	273529	265651	39713	99138	11	7491	9032162	422484	4447442	3781719
2022-03-30	177716	273608	265662	40024	99138	11	7491	9035127	422498	4507681	3807856
2022-03-31	177747	273677	265671	40024	99169	11	7491	9037911	422519	4573066	3842048
2022-04-01	177782	273759	265679	40024	99194	11	7493	9039838	422540	4628163	3868236
2022-04-02	177803	273823	265684	40024	99194	11	7493	9040640	422563	4676260	3890531
2022-04-03	177827	273870	265691	40024	99194	11	7493	9041124	422574	4716870	3909523
2022-04-04	177897	273913	265694	40024	99194	11	7493	9043098	422581	4772414	3923533
2022-04-05	177932	274000	265699	40024	99194	11	7493	9045326	422594	4848211	3935083
2022-04-06	177974	274055	265705	40024	99194	11	7493	9047408	422610	4910819	3950434
2022-04-07	177974	274108	265707	40328	99194	11	7511	9049250	422629	4971300	3971481
2022-04-08	177974	274136	265714	40328	99194	11	7511	9051243	422643	5020617	3987872
2022-04-09	177974	274191	265720	40328	99194	11	7511	9052083	422662	5059642	4001802
2022-04-10	177974	274219	265724	40328	99194	11	7511	9052536	422677	5108366	4014829
2022-04-11	178141	274219	265727	40328	99194	11	7511	9054126	422678	5154633	4025499
2022-04-12	178257	274272	265730	40328	99194	11	7523	9056203	422691	5207650	4032839
2022-04-13	178295	274320	265731	40709	99194	11	7523	9057923	422711	5262359	4042363
2022-04-14	178352	274376	265733	40709	99194	11	7535	9059351	422721	5308858	4054989
2022-04-15	178373	274429	265738	40709	99194	11	7535	9059944	422729	5345438	4065438
2022-04-16	178387	274462	265739	40709	99194	11	7535	9060495	422747	5384615	4075415
2022-04-17	178418	274504	265739	40709	99194	11	7539	9060923	422762	5416980	4084222
2022-04-18	178457	274520	265741	40709	99194	11	7539	9060923	422770	5459012	4091716
2022-04-19	178513	274535	265746	40709	99287	11	7539	9060923	422784	5509592	4097345
2022-04-20	178574	274606	265746	41013	99287	11	7567	9060923	422799	5563493	4103119

### Normalizacja danych.

Określenie dynamiki rozwoju pandemii w danym przedziale czasowym i porównanie z innymi krajami wymaga przekształcenia danych w taki sposób, aby dane były porównywalne i mierzalne bez względu na wielkość populacji danego kraju. Przyjętym rozwiązaniem w tej metodzie jest normalizacja. Dane macierzy  $Conf$  z każdej kolumny z osobna są normalizowane w przedziałach  $\langle 0,1 \rangle$  za pomocą wzoru:

$$Conf_{norm_{col}}(i) = \frac{Conf_{col}(i) - \min(Conf_{col})}{\max(Conf_{col}) - \min(Conf_{col})} \quad (39),$$

gdzie:

$Conf_{norm_{col}}(i)$  – znormalizowana wartość i-tego elementu szeregu czasowego dla danego kraju – kolumny,

$Conf_{col}(i)$  – wartość i-tego elementu (wiersza) w kolumnie - kraju

$\max(Conf_{col})$  – maksymalna wartość dla danej kolumny – kraju, oznaczająca maksymalne skumulowane potwierdzone zakażenia zawierające się w segmencie danych

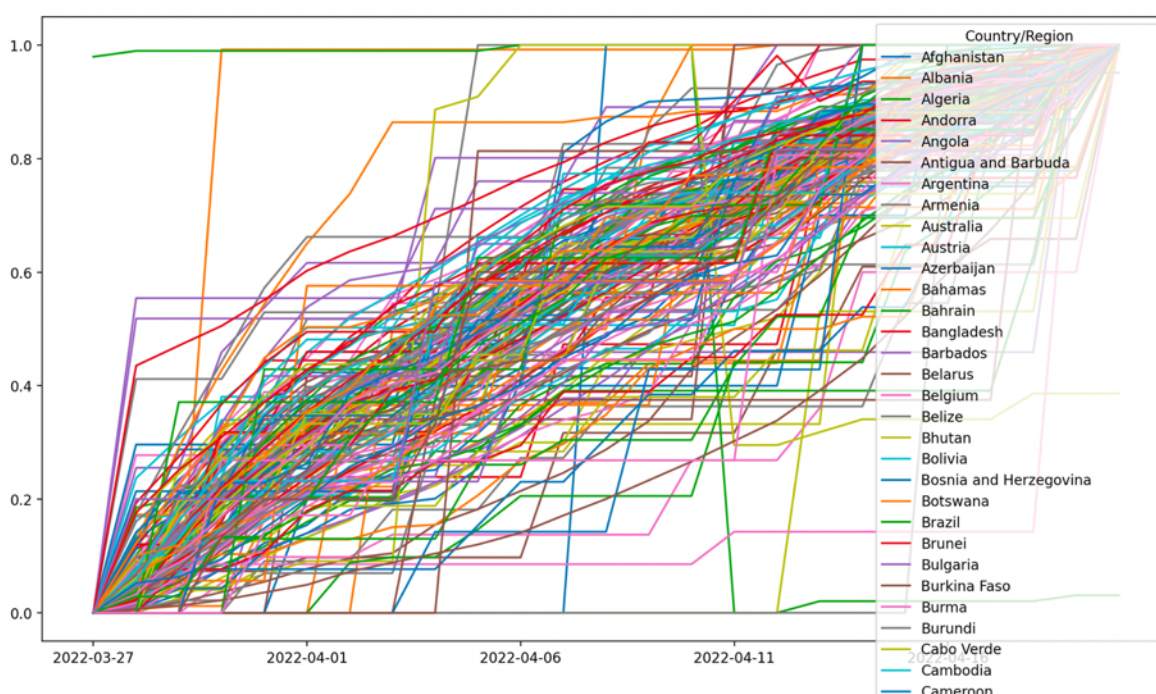
$\min(Conf_{col})$  – minimalna wartość dla danej kolumny – kraju, oznaczająca minimalne skumulowane potwierdzone zakażenia zawierające się w segmencie danych

Efektom normalizacji każdej kolumny z osobna jest macierz  $Conf_{norm}$  (Tab. 15. ).



Otrzymane wartości kolumn macierzy  $Conf_{norm}$  pokazują dynamikę zakażeń w krajach bez względu na wielkość populacji danego kraju. Ich przebieg w badanym oknie zawiera się w przedziale wartości  $<0,1>$  (Rys. 26. ).

W przypadku wystąpienia wartości nie liczbowych dla danego kraju, kraj ten jest usuwany z listy. Nie stosuje się uzupełniania danych, bowiem analiza musi polegać na wiarygodnych i kompletnych danych.



Rys. 26. Otrzymane wartości kolumn macierzy  $Conf_{norm}$ . Przebiegi dla poszczególnych krajów. Źródło: opracowania własne.

Z macierzy  $Conf_{norm}$  przenosimy kolumnę zawierającą kraj bazowy i tworzymy z niej wektor  $Conf_{base}$ , który będzie wykorzystany do walidacji modelu.

### 2.3.2 Analiza zbioru danych wejściowych

Celem jest znalezienie w  $Conf_{norm}$  n-przebiegów zbliżonych do wybranego kraju – określonego jako bazowego  $Conf_{base}$ . Przyjęto tutaj, że kryterium porównawczym będzie suma bezwzględnej wartości odchylenia wartości danego kraju. Kryterium wyboru to n-krajów o

najniższej wartości kryterium porównawczego. Założyć należy, że wśród danych mogą wystąpić przypadki, w których kryterium wyboru będzie zachowane, ale dane będą miały nietypowe przebiegi.

Do wyznaczenia n-najbliższych sąsiadów przyjęto odległość wartości skumulowanych dobowych przyrostów krajów względem kraju bazowego. Przyjęto, że należy uwzględnić dystans w całym segmencie danych, czyli 24 dni. Oparto to na założeniu, że dynamika rozwoju zakażeń jest wypadkową niezliczonych czynników, takich jak poziom szczepień, położenie geograficzne, strefa klimatyczna, ustrój polityczny, stosowanie ograniczeń oraz ich przestrzeganie przez społeczeństwo, itp. Niestety, dane uwzględniające powyższe czynniki nie są raportowane w sposób harmonizowany, o ile w ogóle część z nich jest raportowana. Zatem w niniejszej metodzie opieramy się wyłącznie na danych z CSSE.

Dla każdego dnia – elementu szeregu czasowego – należy wykonać operację odejmowania w celu określenia dziennego dystansu od kraju bazowego:

$$Dist(i) = |Conf_{base}(i) - Conf_{norm}(i)| \quad (40)$$

Wynikiem jest macierz  $Dist$  o wymiarach identycznych jak macierz  $Conf_{norm}$  (Tab. 16. ) zawierająca bezwzględne wartości odległości od przebiegu kraju bazowego dla każdego dnia okna czasowego.



Pierwszy i ostatni wiersz macierzy zawiera elementy o wartości 0 ze względu na to, że dane były znormalizowane w przedziale  $\langle 0,1 \rangle$  i dla każdego kraju są takie same, zaczynające się od 0 i kończące na wartości 1.

Następnie wartości każdej kolumny (zawierającej dane określonego kraju) są sumowane i zapisane w tabeli  $Dist_{sum}$  (Tab. 17. ).

$$Dist_{sum}(C) = \sum_{n=1}^{n=w} Dist_C(i) \quad (41),$$

gdzie:

$Dist_{sum}(C)$ - suma odległości elementów szeregu czasowego kraju C od wartości elementów szeregu czasowego kraju bazowego,

$n$  – indeks elementu w szeregu czasowym

$w$  – rozmiar okna mierzony w dniach

$Dist_C(i)$  – odległość szeregu czasowego kraju C od szeregu kraju bazowego w  $i$ -tym dniu.

Elementy tabeli  $Dist_{sum}$  są następnie sortowane rosnąco. Im mniejsza wartość dla danego kraju, tym bliższe jest jego położenie względem przebiegu kraju bazowego. Z tabeli wybieramy  $n$ -pierwszych wierszy – krajów. o podobnym przebiegu do kraju bazowego (Tab. 18. ).

Tab. 17. Tabela  $Dist_{sum}$  zawierająca odległości przebiegu szeregu czasowego danego kraju od przebiegu szeregu czasowego kraju bazowego. Źródło: opracowanie własne.

Country/Region	0
Monaco	0.395756
France	0.415771
Morocco	0.478906
Australia	0.483653
Colombia	0.49877
Belarus	0.511732
Albania	0.516904
Kosovo	0.521085
Bahrain	0.559753
Thailand	0.57555
Philippines	0.583588
Moldova	0.589857

W tym przypadku n-najbliższymi krajami są:

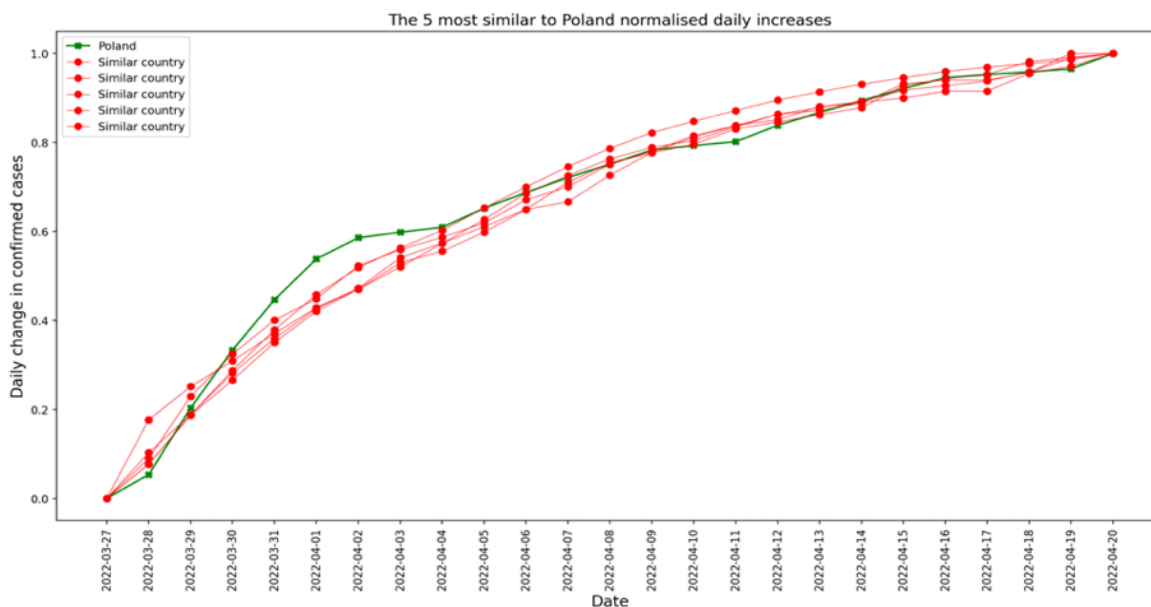
Tab. 18.  $N=5$  najbliższych krajów i ich odległości od kraju bazowego  $Dist_{sum}(C)$ . Źródło: opracowania własne.

Kraj C	$asc(Dist_{sum}(C))$
Burma	0,55
Sri Lanka	0,78
Vietnam	0,80
Netherlands	0,81
Algeria	0,88

### 2.3.3 Walidacja i weryfikacja wyniku

Istotnym krokiem jest walidacja i weryfikacja dokonanego wyboru n-krajów. Dokonuje się tego za pomocą analizy statystycznej (Tab. 19. ) oraz graficznej inspekcji danych.

Pierwszym krokiem jest wygenerowanie i porównanie przebiegów szeregów czasowych badanego kraju oraz jego n-najbliższych sąsiadów. Przykładowy wykres przedstawiono na Rys. 27.



Rys. 27. Wybrane kNN-najbliższe przebiegi względem kraju bazowego (oznaczone kolorem czerwonym) oraz przebieg kraju bazowego (oznaczony kolorem zielonym). Źródło: opracowanie własne.

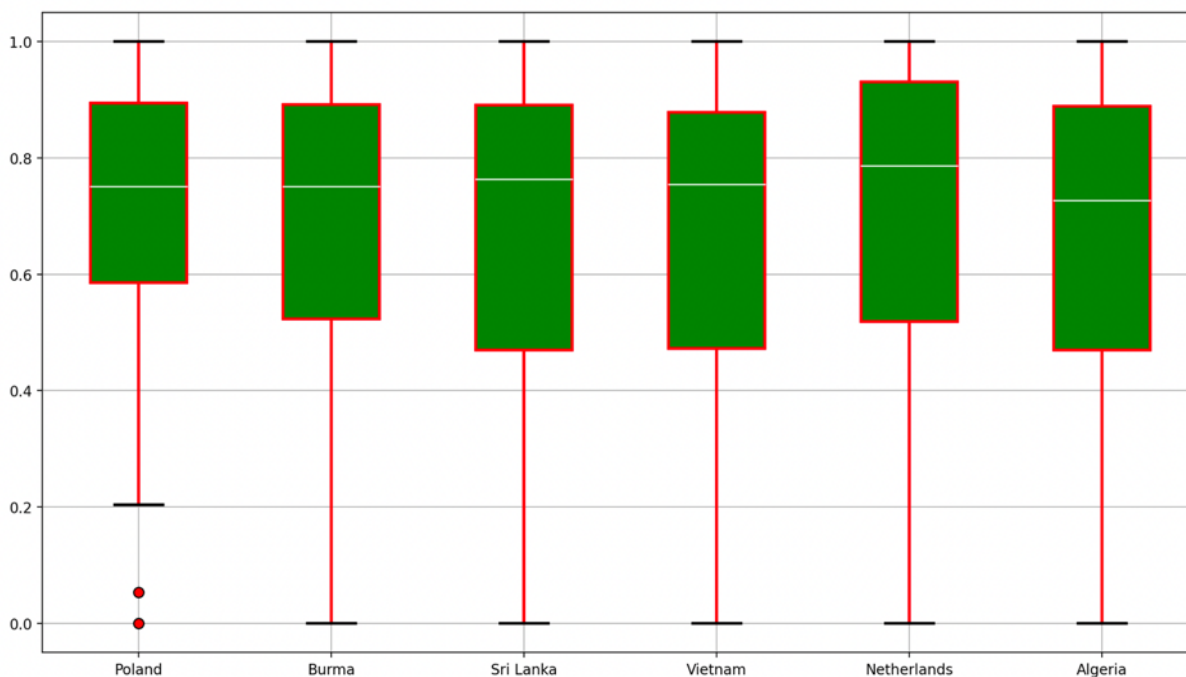
Należy sprawdzić wśród przebiegów szeregu czasowego  $Conf_{norm}$  w kolumnach n-najbliższych krajów występowanie wartości o relatywnie ekstremalnej wartości odchylenia w stosunku do pozostałych wartości przebiegu, czyli tzw - ‘outlierów’. Zastosowano w tej metodzie techniki:

1. Analiza wartości standardowego odchylenia kolumny dla danego kraju: jeśli odchylenie standardowe ma zbliżone wartości dla zbioru n-krajów, nie powinno być outlierów.
2. Analiza graficzna w postaci wykresu pudełkowego. W tym przypadku zestawia się dane z kraju bazowego oraz n-najbliższych sąsiadów, bowiem może okazać się, że ‘outlier’ będzie obecny w danych kraju bazowego.

Tab. 19. *Analiza statystyczna segmentu danych dla n-najbliższych krajów. Kolumna ‘std’ zawiera wartość średniego odchylenia elementów szeregu czasowego. Źródło: opracowania własne.*

Country/Region	count	mean	std	min	25%	50%	75%	max
Burma	25	0.667708	0.28532	0	0.522815	0.750962	0.891699	1
Sri Lanka	25	0.655369	0.294748	0	0.470227	0.763233	0.890832	1
Vietnam	25	0.665443	0.282634	0	0.472672	0.754009	0.87795	1
Netherlands	25	0.683135	0.302128	0	0.518666	0.786754	0.930889	1
Algeria	25	0.655043	0.294798	0	0.470085	0.726496	0.888889	1

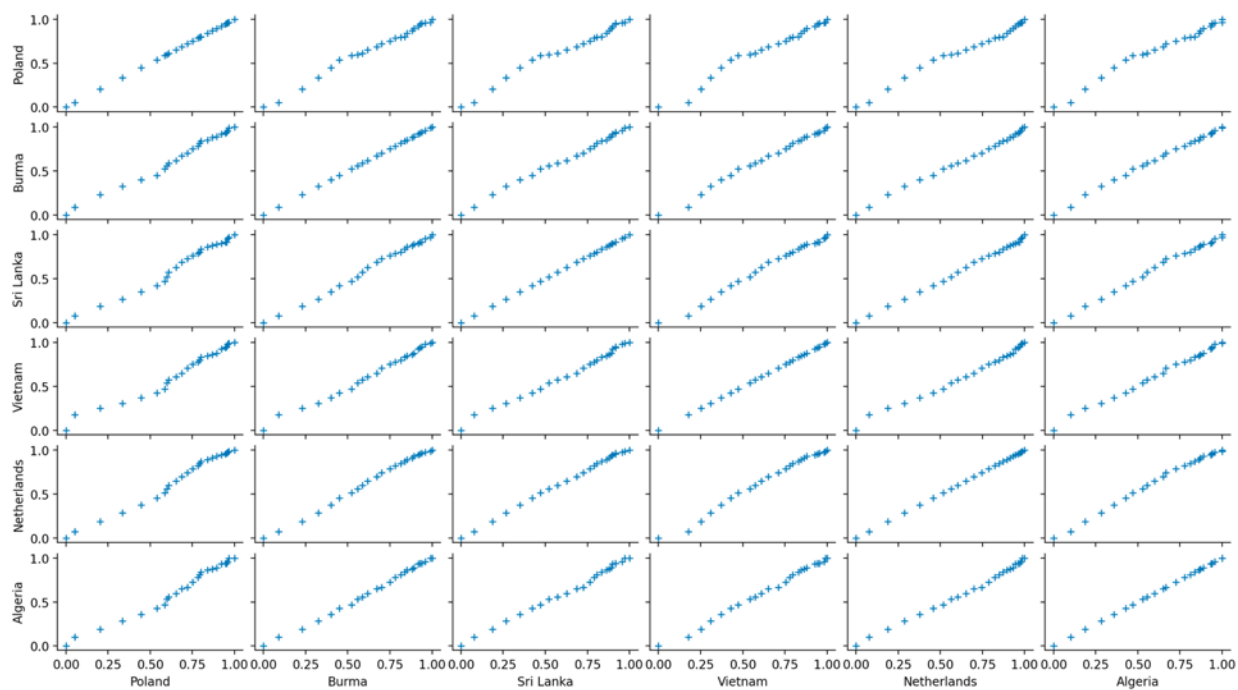
Dodatkowo zobrazujemy na wykresie pudełkowym na Rys. 28. dane n-najbliższych krajów – kolumn z macierzy Conf:



Rys. 28. Wykres pudełkowy przedstawiający w sposób graficzny reprezentację podstawowej analizy statystycznej kraju bazowego (tutaj Polska) oraz  $n$ -najbliższych sąsiadów. Pierwszy kraj to kraj bazowy, kolejne to jego najbliżsi sąsiedzi. Na przykładzie kraju bazowego widać 2 outliery reprezentowane w postaci czerwonych punktów.

Badanie zależności pomiędzy  $n$ -najbliższymi sąsiadami ma na celu weryfikację założenia, że dane  $n$ -najbliższych sąsiadów i kraju bazowego są ze sobą powiązane. W tym celu dokonuje się graficznej inspekcji na wykresie dwuwymiarowym par krajów oraz badany jest współczynnik korelacji Pearson'a.

Kolejnym krokiem badającym zależność przebiegu dla kraju bazowego i  $n$ -najbliższych sąsiadów jest sprawdzenie zależności rozkładów na wykresach dwuwymiarowych (Rys. 29. ). Wykres taki przedstawia zależność dla  $(n,2)$  kombinacji zmiennych w zbiorze analizowanych danych jako macierz wykresów. Jeśli istnieje relacja pomiędzy dwoma krajami w ich przebiegach, wartości przedstawione na wykresie będą zbieżne ze sobą i tworzyły linię krzywą. W przypadku braku zbieżności będą to chaotycznie porzucane punkty na wykresie.



Rys. 29. Wykres rozkładów dwuwymiarowych par w zbiorze danych reprezentujących kraj bazowy i  $n$ -najbliższych sąsiadów.

Korelacja Pearson'a pomimo swoich niedoskonałości w przypadku małych zbiorów danych pozwala określić zależność pomiędzy zmiennymi (Steuer et. Al. (2002)). Następnym krokiem jest obliczenie i zobrazowanie korelacji Pearson'a kraju bazowego i  $n$ -najbliższych krajów (Rys. 30. ). Im bliższa wartość do 1 tym korelacja jest większa.



Rys. 30. Mapa ciepła obrazująca korelację wzajemną między krajem bazowym oraz  $n$ -najbliższymi sąsiadami. Źródło: opracowanie własne.

### 2.3.4 Model oparty na średniej kroczącej.

Model predykcji bazuje na średnich kroczących  $n$ -najbliższych sąsiadów oraz krzywej utworzonej na podstawie ich uśrednionych dziennych wartości. Będzie to krzywa referencyjna, której przedłużenie w oknie predykcji będzie zawierało wartości predykcji dla kraju bazowego.

Obliczenie średniej kroczącej dla przebiegów  $n$ -najbliższych sąsiadów dokonywane jest dla każdego kraju z osobna. Przyjęto, że będzie to 7-dniowa średnia krocząca  $MA_7$  obliczana zgodnie ze wzorem:

$$MA_7 = \frac{\sum_{i=1}^{i=w} Conf_{norm}^C(i)}{7} \quad (42),$$

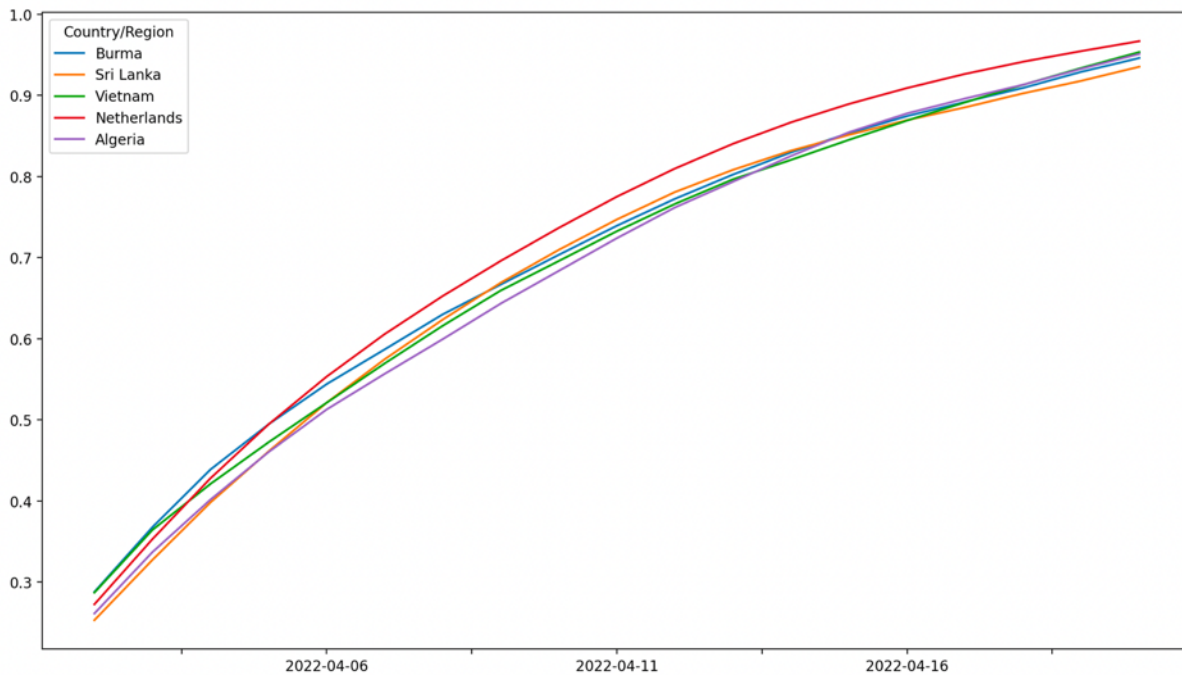
gdzie:

$C(i)$  – wartość skumulowanych dobowych przyrostów zawartych w macierzy  $Conf_{norm}$  dla kraju  $C$  w  $i$ -tym dniu 7 dniowego okna  $w$ .

Wyniki umieszczone są w wierszu oznaczającym ostatni dzień w oknie w. Stąd w tabeli Tab. 20. z wynikami brak jest wartości liczbowych w pierwszych 6 wierszach. Dane z Tab. 20. zostały przedstawione w postaci wykresu na Rys. 31.

Tab. 20. Wartości średniej kroczącej o rozmiarze okna 7 dni dla n-sąsiadujących krajów.  
 Źródło: opracowania własne.

Index	Burma	Sri Lanka	Vietnam	Netherlands	Algeria
2022-03-27	nan	nan	nan	nan	nan
2022-03-28	nan	nan	nan	nan	nan
2022-03-29	nan	nan	nan	nan	nan
2022-03-30	nan	nan	nan	nan	nan
2022-03-31	nan	nan	nan	nan	nan
2022-04-01	nan	nan	nan	nan	nan
2022-04-02	0.287913	0.253038	0.287063	0.272582	0.261294
2022-04-03	0.367784	0.32737	0.364364	0.352995	0.336996
2022-04-04	0.43886	0.398258	0.421099	0.428046	0.401709
2022-04-05	0.494542	0.461113	0.472408	0.494451	0.460317
2022-04-06	0.544098	0.521132	0.521097	0.553323	0.512821
2022-04-07	0.5869	0.574737	0.569583	0.605769	0.556777
2022-04-08	0.630095	0.623616	0.616061	0.652721	0.599512
2022-04-09	0.666928	0.66912	0.659522	0.696077	0.643468
2022-04-10	0.703526	0.709695	0.69592	0.736779	0.683761
2022-04-11	0.739417	0.747097	0.732624	0.775232	0.724054
2022-04-12	0.772559	0.780921	0.766245	0.809908	0.761905
2022-04-13	0.802482	0.808331	0.796627	0.840471	0.793651
2022-04-14	0.829812	0.831961	0.820569	0.866898	0.825397
2022-04-15	0.853609	0.851472	0.845358	0.889602	0.854701
2022-04-16	0.874578	0.869565	0.869114	0.909209	0.8779
2022-04-17	0.89217	0.885431	0.891562	0.9266	0.896215
2022-04-18	0.909369	0.902647	0.913182	0.941676	0.913309
2022-04-19	0.92916	0.918039	0.93391	0.954727	0.932845
2022-04-20	0.946203	0.935458	0.95355	0.96704	0.95116



Rys. 31. Wykres przebiegu średniej kroczącej  $MA_7$  dla każdego z  $n$ -najbliższych sąsiadów dla kraju bazowego. Źródło: opracowanie własne.

Predykcja będzie bazować na **krzywej referencyjnej** i jej przedłużeniu w oknie predykcji.

Otrzymane wartości średnich kroczących  $MA_7$  dla każdego z kraju jest uśredniane za pomocą średniej arytmetycznej zgodnie ze wzorem:

$$MA_{avg}(i) = \sum_{k=1}^{k=n} MA_k(i) \quad (43),$$

gdzie:

$MA_{avg}$  - to średnia arytmetyczna wartości  $MA_7$  w kończąca się w oknie w  $i$ -tym dniu

$k \in \langle 1, \dots, n \rangle$  -  $k$  kolejny kraj z  $n$ -najbliższych sąsiadów

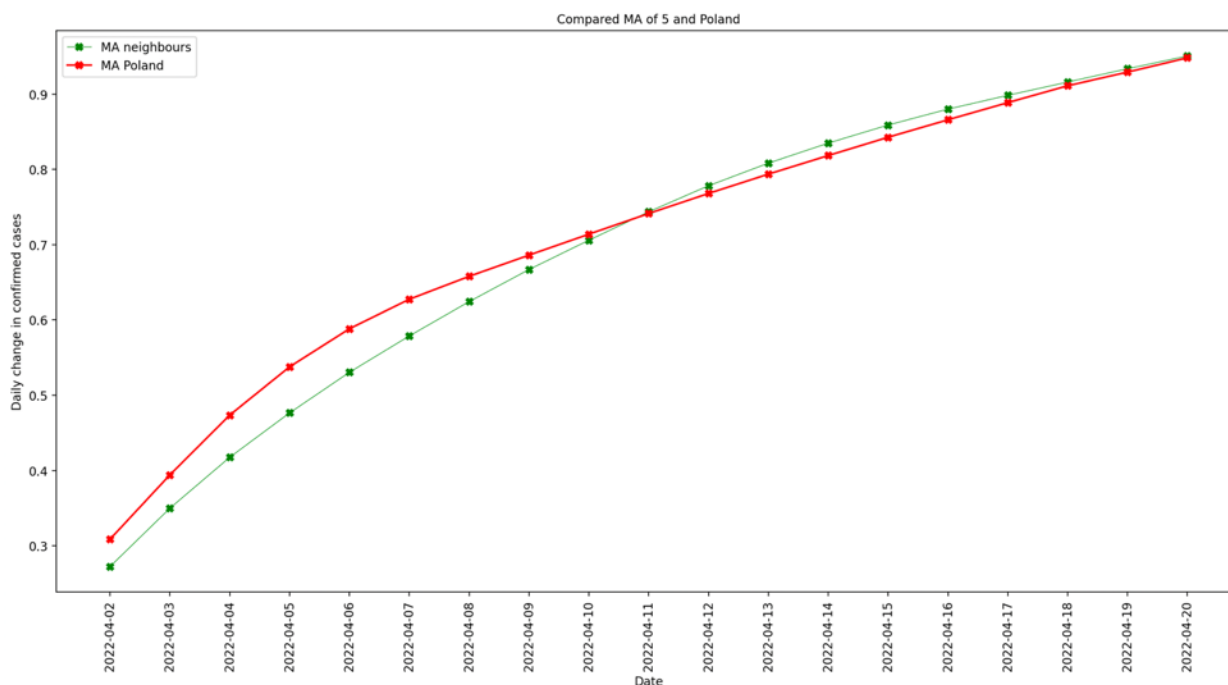
$MA_k$  - wartość średniej kroczącej w oknie w kończącym się w  $i$ -tym dniu dla  $k$  kolejnego kraju z  $n$ -najbliższych sąsiadów.

Wynikiem obliczeń jest kolumna 'avg' z Tab. 21. reprezentująca szereg czasowy tworzący przebieg krzywej referencyjnej zobrazowany na Rys. 32.

Tab. 21. Przykład uśrednionych wartości dziowych umieszczonych w kolumnie 'avg'.  
Źródło: opracowanie własne.

Index	Burma	Sri Lanka	Vietnam	Netherlands	Algeria	avg
2022-03-27	nan	nan	nan	nan	nan	nan
2022-03-28	nan	nan	nan	nan	nan	nan
2022-03-29	nan	nan	nan	nan	nan	nan
2022-03-30	nan	nan	nan	nan	nan	nan
2022-03-31	nan	nan	nan	nan	nan	nan
2022-04-01	nan	nan	nan	nan	nan	nan
2022-04-02	0.287913	0.253038	0.287063	0.272582	0.261294	0.272378
2022-04-03	0.367784	0.32737	0.364364	0.352995	0.336996	0.349902
2022-04-04	0.43886	0.398258	0.421099	0.428046	0.401709	0.417594
2022-04-05	0.494542	0.461113	0.472408	0.494451	0.460317	0.476566
2022-04-06	0.544098	0.521132	0.521097	0.553323	0.512821	0.530494
2022-04-07	0.5869	0.574737	0.569583	0.605769	0.556777	0.578753

Krzywa referencyjna powinna mieć zbliżony przebieg do przebiegu kraju bazowego.



Rys. 32. Krzywa referencyjna (oznaczona kolorem zielonym) zestawiona z średnią kroczącą kraju bazowego (oznaczona kolorem czerwonym) w przedziale uczenia. Źródło: opracowanie własne.

### 2.3.5 Predykcja dobowych skumulowanych zakażeń.

Predykcja w horyzoncie predykcji  $hp$  dni obliczana jest za pomocą następującego wzoru:

$$Y_m(i) = Y_m(i - 1) + [MA_{avg}(i - hp - 1) - MA_{avg}(i - hp)] \quad (44) ,$$

gdzie:

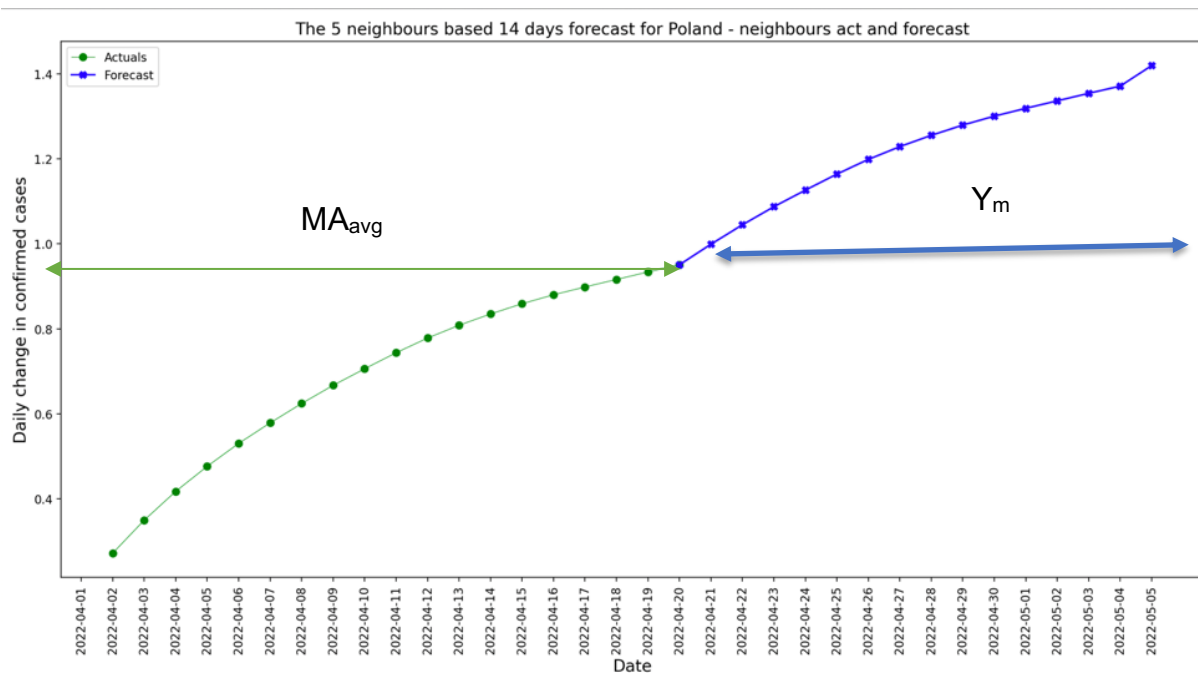
$Y_m(i)$  – wartość przewidywana w  $i$ -tym dniu horyzontu predykcji  $hp$ .

$Y_m(i - 1)$  – wartość predykcji z dnia poprzedzającego. Dla  $i=1$  przyjmuje się wartość  $MA_{avg}(i-1)$ .

$MA_{avg}(i - hp - 1)$  – wartość przebiegu krzywej referencyjnej  $hp+1$  dni przed dniem, dla którego dokonuje się predykcji

$MA_{avg}(i - hp)$  - wartość przebiegu krzywej referencyjnej  $hp$  dni przed dniem, dla którego dokonuje się predykcji

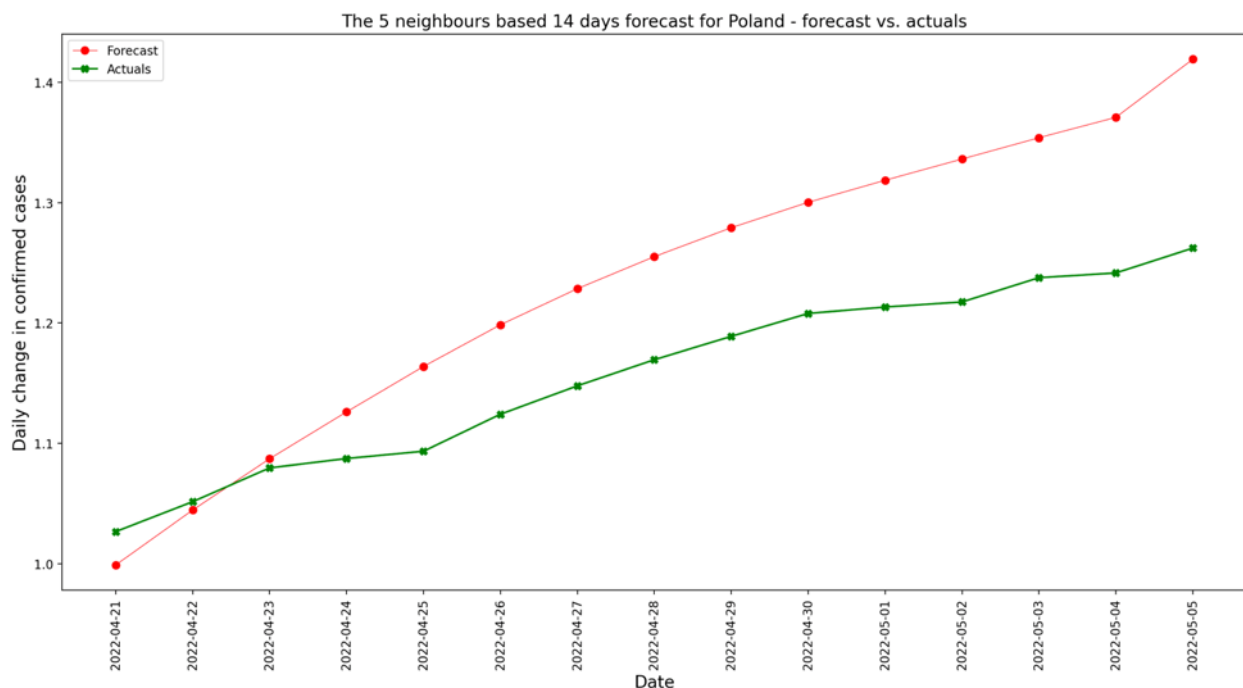
Zobrazowane przebiegi krzywej referencyjnej i przebiegu przewidywanych wartości szeregu czasowego przedstawiono na Rys. 33.



Rys. 33. Zobrazowane przebiegi krzywej referencyjnej i przebiegu przewidywanych wartości szeregu czasowego. Źródło: opracowanie własne.

### 2.3.6 Ewaluacja wyników

Wyniki predykcji analizuje się na podstawie MAPE (Tab. 22. ). Zalecana jest również analiza graficzna (Rys. 34. ). Przebieg prognozowanych wartości oraz wartości rzeczywistych dla kraju bazowego porównuje się graficznie poprzez zestawienie obydwu krzywych. Im mniejszy dystans krzywych od siebie, tym wynik powinien być lepszy.

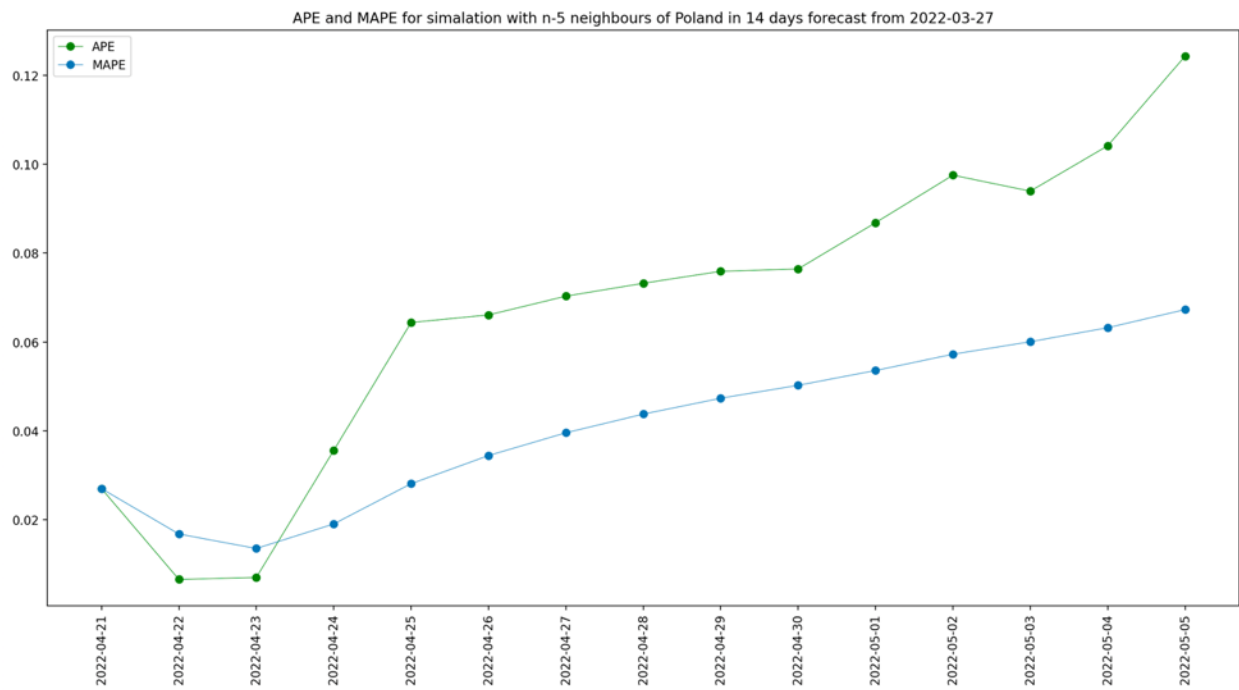


Rys. 34. *Przebieg wartości rzeczywistych (kolor zielony) oraz prognozowanych (kolor czerwony) dla kraju bazowego w horyzoncie predykcji  $h_p=14$  dni. Źródło: opracowanie własne.*

Następnie obliczany jest MAPE oraz APE. Im mniejsze wartości, tym lepiej. Wykresy MAPE i APE dla kolejnych dni prognozy przedstawiono na Rys. 35.

Tab. 22. Wynik symulacji zawierający prognozowane wartości (kolumna 'avg'), przebieg rzeczywisty (kolumna 'actual\_cases\_no'), rzeczywista liczba skumulowanych wartości przyrostów dobowych (kolumna 'actual\_cases'), APE (kolumna 'ape') oraz MAPE (kolumna 'MAPE'). Źródło: opracowanie własne.

Index	avg	tual_cases_no	actual_cases	ape	MAPE
2022-04-21	0.998941	1.02668	5988518	0.0270189	0.0270189
2022-04-22	1.04459	1.05153	5989614	0.00659687	0.0168079
2022-04-23	1.08721	1.07961	5990853	0.00703877	0.0135515
2022-04-24	1.12612	1.08741	5991197	0.0356023	0.0190642
2022-04-25	1.16387	1.09346	5991464	0.0643922	0.0281298
2022-04-26	1.1985	1.1242	5992820	0.0660866	0.0344559
2022-04-27	1.2285	1.1478	5993861	0.0703098	0.0395779
2022-04-28	1.25512	1.16949	5994818	0.0732136	0.0437824
2022-04-29	1.27914	1.1889	5995674	0.0759018	0.0473512
2022-04-30	1.30026	1.20794	5996514	0.0764299	0.0502591
2022-05-01	1.31858	1.21324	5996748	0.0868256	0.0535833
2022-05-02	1.33622	1.21748	5996935	0.0975313	0.0572456
2022-05-03	1.35392	1.23766	5997825	0.0939416	0.0600684
2022-05-04	1.37087	1.24158	5997998	0.104135	0.063216
2022-05-05	1.41913	1.26223	5998909	0.124303	0.0672885



Rys. 35. Wykres przebiegów APE (linia koloru zielonego) oraz MAPE (linia koloru niebieskiego) w horyzoncie predykcji hp. Źródło: opracowanie własne.

## 2.4 Prognoza inflacji na poziomie kraju.

### 2.4.1 Metoda regresji wielomianem z optymalizacją parametrów.

W regresji wielomianowej model matematyczny stara się dopasować wielomian  $n$ -tego stopnia do danych, co pozwala przewidzieć przyszłe wartości na podstawie historycznych obserwacji. Regresja wielomianowa jest metodą analizy regresji, w której zależność między zmienną zależną  $Y$  a zmienną niezależną  $X$  (lub zmiennymi) jest modelowana jako wielomian  $n$ -tego stopnia. Model matematyczny wyraża się następującym wzorem:

$$Y = a_0 + a_1X + a_2X^2 + \dots + a_nX^n + \epsilon \quad (45)$$

gdzie:

- $Y$  - zmienna zależna, reprezentująca wartości szeregu czasowego inflacji
- $X$  - zmienna niezależna, reprezentująca szereg czasowy inflacji dla danego kraju.
- $a_0, a_1, \dots, a_n$ ) współczynniki wielomianu.

- $n$  - stopień wielomianu.
- $\epsilon$  - termin błędu, uwzględniający niedokładność modelu.

Dane wejściowe szeregu czasowego są interpretowane jako zmienne niezależne  $X$  w modelu regresji. Każdy punkt w szeregu czasowym jest traktowany jako osobna obserwacja z odpowiednią wartością czasu (lub innego indeksu porządkowego) i przypisaną do niej wartością zależną  $Y$ . W kontekście prognozowania, dane historyczne są wykorzystywane do trenowania modelu, a następnie model jest wykorzystywany do generowania przewidywań na przyszłe momenty czasowe.

W kontekście zastosowanego modelowania szeregu czasowego  $hp$  oznacza horyzont predykcji,  $kwst$  to liczba kroków wstecz (rozmiar okna uczącego), a  $n$  to stopień wielomianu. Model sformułowany jest w następujący sposób:

$$Y_{\{t+1\}} = a_0 + a_1 X_{\{t-(kwst-1)\}} + a_2 X_{\{t-(kwst-1)\}}^2 + \dots + a_n X_{\{t-(kwst-1)\}}^n + \epsilon \quad (46)$$

gdzie:

$Y_{\{t+1\}}$  – wartość będąca rezultatem predykcji

$a_0, a_1, \dots, a_n$  – szukane współczynniki wielomianu

$X_{t-(kwst-1)}$  - wartość zmiennej niezależnej odpowiadająca punktowi czasowemu  $t - (kwst - 1)$

$\epsilon$  - błąd modelu, reprezentujący różnicę między modelem a rzeczywistymi danymi.

Dla każdego punktu w oknie uczenia  $kwst$  otrzymujemy równanie:

$$Y_{\{t+1\}} = a_0 + a_1(t - (kwst - 1)) + a_2(t - (kwst - 1))^2 + \dots + a_n(t - (kwst - 1))^n + \epsilon. \quad (47)$$

Równanie (47) jest powtarzane dla każdego punktu w oknie  $kwst$ , co oznacza, że dla  $kwst$  punktów czasowych, otrzymujemy  $kwst$  równań, które są następnie wykorzystywane do znalezienia najlepszych współczynników  $a_0, a_1, \dots, a_n$  w sensie minimalizacji błędu  $\epsilon$  przy użyciu metody najmniejszych kwadratów.

Zaimplementowano model z optymalizacją parametrów  $kwst$  i  $n$ . Algorytm metodą prób i błędów szuka najlepszego modelu predykcji dla szeregu czasowego (dającego najniższy błąd MSE) poprzez iteracje odpowiadające kombinacji po parametrach  $kwst$  i  $n$  w zakresie wartości

całkowitych odpowiednio 2-25 i 1-25. Poszukiwanie optymalnych parametrów i predykcja odbywają się w następujących krokach:

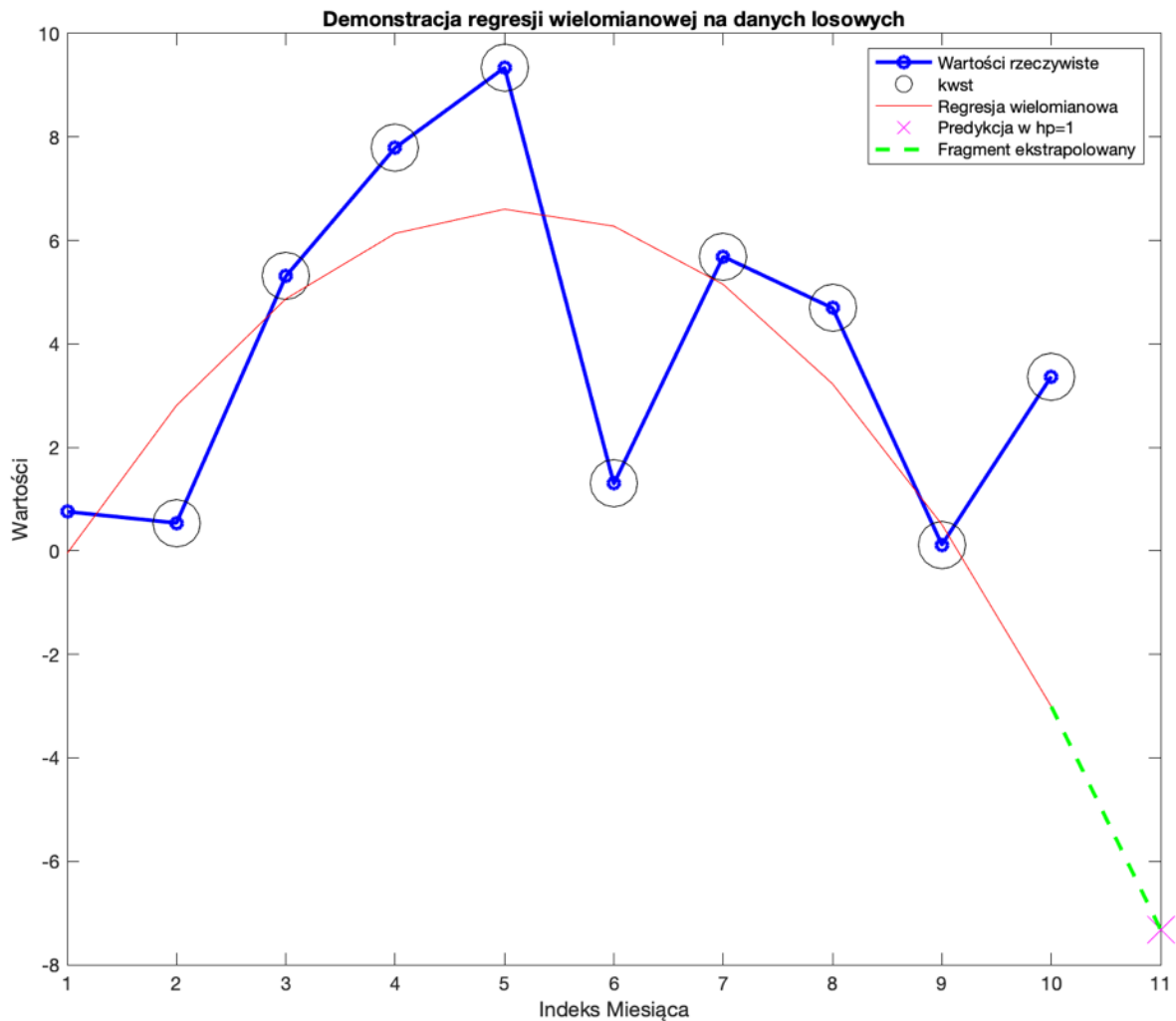
1. Przed przystąpieniem do dopasowania wielomianu, kod sprawdza, czy istnieje wystarczająca liczba punktów danych ( $\text{length}(\mathbf{Y}_{\text{train}}) \geq \text{kwst}$ ), aby przeprowadzić regresję dla danego rozmiaru okna.
2. Jeśli jest wystarczająca liczba danych, kod wykorzystuje funkcję **polyfit** do dopasowania wielomianu stopnia **n** do danych **Yr**, które są ostatnimi **kwst** wartościami z zestawu treningowego **Y\_train**.
3. Następnie, używając funkcji **polyval**, kod przewiduje wartość (**Yp**) w następnym kroku czasowym ( $\text{length}(\mathbf{Yr}) + 1$ ), co odpowiada wartości **Yn(i+1)**.
4. Dla każdej kombinacji parametrów **n** i **kwst**, kod oblicza błąd średniokwadratowy (MSE) między przewidywaną wartością (**Yp**) a rzeczywistą wartością w następnym kroku czasowym (**Yn(i+1)**). Jeżeli obliczony MSE jest niższy niż dotychczas znaleziony minimalny MSE (**min\_mse**), kod aktualizuje zmienne **min\_mse** i **optimal\_params** nowymi, lepszymi parametrami. Po przejściu przez wszystkie możliwe kombinacje **n** i **kwst**, kod zakończy posiadając optymalne parametry dla modelu, które minimalizują MSE i teoretycznie powinny zapewnić najlepszą możliwą predykcję dla danych szeregu czasowego.
5. Następnie za pomocą optymalnych parametrów znalezionych w poprzednim kroku dokonuje się predykcji na następny miesiąc. Wykorzystuje dwie kluczowe funkcje MATLAB-a: **polyfit** („Polynomial curve fitting - MATLAB polyfit”, [https://www.mathworks.com/help/matlab/ref/polyfit.html?searchHighlight=polyfit&s\\_tid=srchtitle\\_support\\_results\\_1\\_polyfit, 2023](https://www.mathworks.com/help/matlab/ref/polyfit.html?searchHighlight=polyfit&s_tid=srchtitle_support_results_1_polyfit, 2023)) i **polyval** („Polynomial evaluation - MATLAB polyval”, [https://www.mathworks.com/help/matlab/ref/polyval.html?s\\_tid=doc\\_ta, 2023](https://www.mathworks.com/help/matlab/ref/polyval.html?s_tid=doc_ta, 2023)).

Funkcja **polyfit** jest używana do dopasowania wielomianu do zestawu danych, tak jak przedstawiono na Rys. 36. Działa poprzez minimalizację sumy kwadratów różnic między danymi a wartościami przewidywanymi przez wielomian. W tym przypadku:

```
optimal_Y_train = Y_train(end-optimal_params(2)+1:end);
```

```
p_full = polyfit(1:length(optimal_Y_train), optimal_Y_train,  
optimal_params(1));  
Yp_next = polyval(p_full, length(optimal_Y_train) + hp);
```

W powyższym kodzie MATLAB, dopasowanie modelu wielomianowego stopnia określonego przez  $n=\text{optimal\_params}(1)$  odbywa się na podstawie ostatnich  $kwst=\text{optimal\_params}(2)$  punktów danych ( $\text{optimal\_Y\_train}$ ). Wynikowe współczynniki wielomianu są następnie używane do przewidywania wartości w kolejnym punkcie czasowym ( $\mathbf{Yp\_next}$ ) za pomocą funkcji **polyval**. Funkcja ta oblicza wartość wielomianu w punkcie określonym jako długość wyselekcjonowanego okna danych plus jeden, co odpowiada predykcji na następny krok czasowy.



Rys. 36. Zobrazowanie ekstrapolacji regresji wielomianowej z uwzględnieniem parametru kwst – kroków wstecz. Niebieskie punkty połączone niebieską linią oznaczają wartości rzeczywiste. Niebieskie kółka oznaczają liczbę kroków wstecz (wartość parametru kwst). Czerwona linia to to wielomian  $n$ -tego stopnia i jego ekstrapolowana część oznaczona zieloną przerywaną linią do punktu oznaczonego X w kolorze magenta – prognozowana wartość w  $hp=1$ . Źródło: opracowanie własne.

#### 2.4.2 Metoda korekcji błędów predykcji za pomocą nieliniowej regresji.

W analizie szeregów czasowych inflacji często występują nieregularności, które mogą wpływać na dokładność modeli predykcyjnych. Aby rozwiązać ten problem, zastosowano innowacyjną metodę korekcji predykcji, wykorzystując nieliniową regresję do modelowania i korygowania błędów predykcyjnych. Model korekcyjny błędów (ECM od ang. Error Correction

Model) opiera się na zastosowaniu nieliniowej regresji do estymacji błędów (różnic między wartościami rzeczywistymi a przewidywanymi). Błędy te są modelowane jako funkcja nieliniowa czasu, co pozwala na ich precyzyjne wyznaczenie i skorygowanie. Model matematyczny błędów jest reprezentowany przez równanie:

$$\epsilon_{pred_i} = b_1 \cdot t_i^2 + b_2 \cdot t_i + b_3 \quad (48)$$

gdzie:

- $\epsilon_{pred_i}$  - błąd predykcji dla i-tego punktu,
- $b_1, b_2, b_3$  – to parametry modelu nieliniowego,
- $t_i$  - indeks czasowy i-tego punktu.

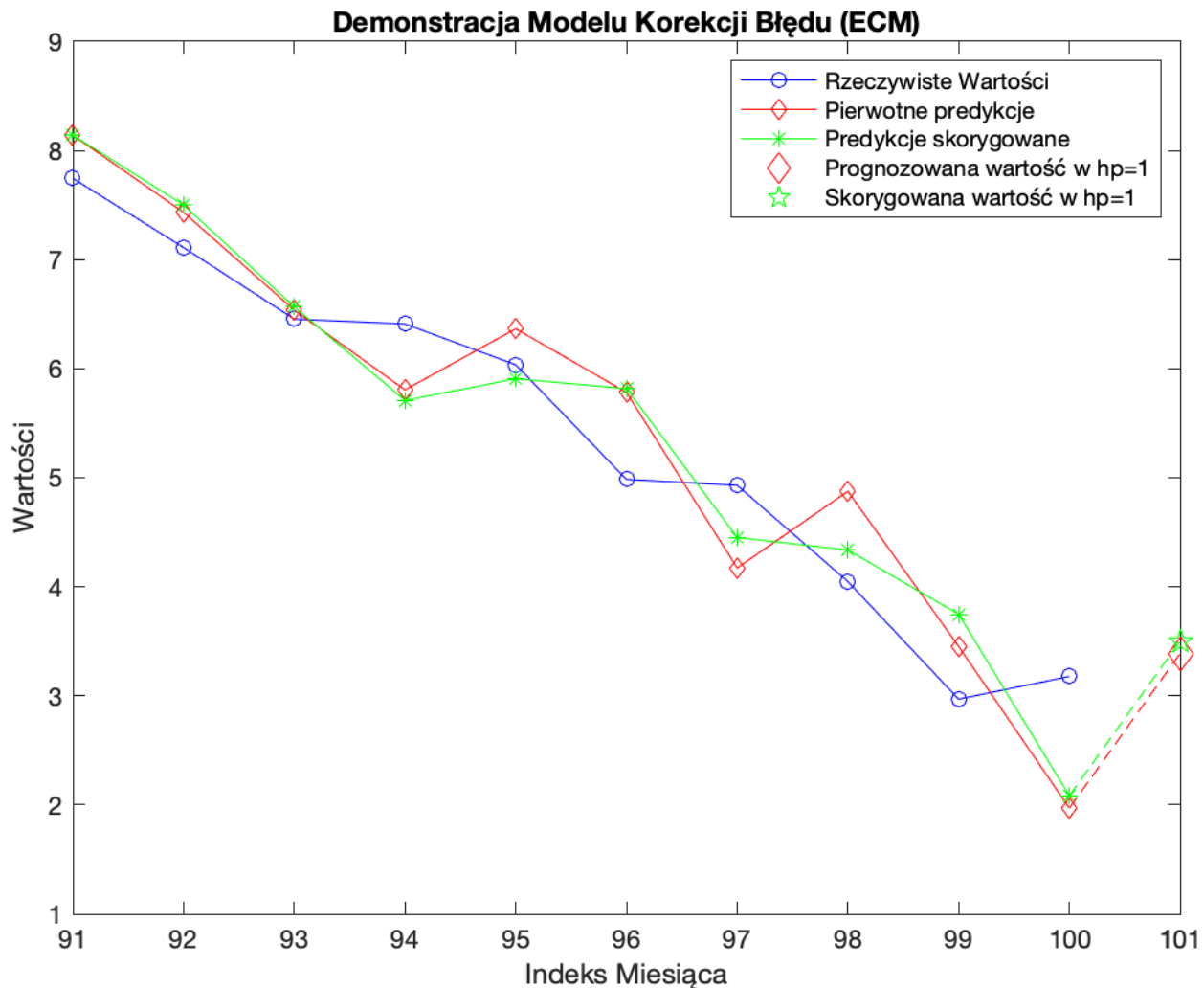
Model nieliniowy jest dopasowywany do obserwowanych błędów predykcji przy użyciu metody najmniejszych kwadratów. W przedstawionym powyżej podejściu jest realizowany przez funkcję **fitnlm** („Fit nonlinear regression model - MATLAB fitnlm” [https://www.mathworks.com/help/stats/fitnlm.html?searchHighlight=fitnlm%20&s\\_tid=srchtitle\\_support\\_results\\_1\\_fitnlm%20 2023](https://www.mathworks.com/help/stats/fitnlm.html?searchHighlight=fitnlm%20&s_tid=srchtitle_support_results_1_fitnlm%20 2023)) w MATLAB, która pozwala na elastyczne dopasowanie modelu do danych. Następnie, na podstawie dopasowanego modelu, obliczane są przewidywane błędy dla każdego punktu w szeregu czasowym. Korekta wartości przewidywanych jest realizowana przez dodanie obliczonego błędu do początkowej wartości przewidywanej, według wzoru:

$$Y_{corr\{t+1\}} = Y_{\{t+1\}} + \epsilon_{pred_i} \quad (49)$$

gdzie:

- $Y_{corr\{t+1\}}$  - skorygowana wartość predycyjna dla i-tego punktu.
- $Y_{\{t+1\}}$  to oryginalna wartość przewidywana dla i-tego punktu otrzymana za pomocą predykcji metodą regresji wielomianem z optymalizacją parametrów.
- $\epsilon_{pred_i}$  to przewidywany błąd modelu dla i-tego punktu.

Zastosowanie nieliniowej regresji do korekcji błędów predykcji w szeregach czasowych (zobrazowana na Rys. 37. ) pozwala na znaczące zwiększenie dokładności modeli predycyjnych. Metoda ta jest szczególnie przydatna w przypadkach, gdzie występują nieregularne fluktuacje, które tradycyjne metody mogą nie uwzględniać.



Rys. 37. *Zobrazowanie graficzne korygowania wartości prognozowanych za pomocą korekty ECM. Niebieska linia oraz punkty oznaczają wartości rzeczywiste, czerwona linia oraz czerwone punkty to pierwotna predykcja – z wcześniejszych kroków, zielona linia i punkty to skorygowane wartości poprzedniej predykcji. Czerwony diament oznacza przewidywaną wartość w  $hp=1$ , natomiast zielona gwiazda oznacza skorygowaną wartość przewidywaną na podstawie błędów z poprzednich kroków. Źródło: opracowanie własne.*

### 2.4.3 Metoda regresji wygładzania splajnem z optymalizacją parametrów.

W przypadku wygładzania splajnem, zastosowano kryterium minimalizacji sumy kwadratów błędów (SQE) między danymi a wartościami estymowanymi przez splajn:

$$SQE = \sum_{j=1}^n [y_j - s(x_j)]^2 \quad (50)$$

gdzie:

$y_j$  - wartości obserwowane,

$S(x_j)$  - wartość estymowana przez splajn w punkcie  $x_j$

$n$  - liczba punktów danych.

W splajnie wygładzającym, wprowadza się parametr wygładzania  $\lambda$ , który kontroluje równowagę między dopasowaniem do danych a gładkością krzywej. Funkcja do minimalizacji może być wtedy zapisana jako:

$$SQE + \lambda \int [S''(x)]^2 dx \quad (51)$$

gdzie:

$\lambda$  – parameter wygładzania,

$\int [S''(x)]^2 dx$  - integralna miara krzywizny splajnu.

Splajn kubiczny na każdym segmencie jest opisany jako:

$$S_{i(x)} = a_i + b_i(x - x_i) + c_i^2(x - x_i) + d_i^3(x - x_i) \quad (52)$$

gdzie  $a_i, b_i, c_i, d_i$  - współczynniki splajnu na segmencie  $i$ -tym.

Aby zapewnić gładkość splajnu, jego pierwsza i druga pochodna muszą być ciągłe na granicach segmentów:

$$S'_{\{i-1\}}(x_i) = S'_i(x_i) \quad (53)$$

$$S''_{\{i-1\}}(x_i) = S''_i(x_i) \quad (54)$$

Do prognozowania przyszłych wartości, splajn jest ekstrapolowany poza zakres danych historycznych:

$$\hat{y}_{\{n+hp\}} = S_{n(x_n+hp)} \quad (55)$$

gdzie  $\hat{y}_{\{n+hp\}}$  to prognozowana wartość na  $hp$  kroków do przodu od ostatniego punktu danych  $x_n$ .

Te wzory stanowią podstawę matematyczną modelu wygładzania splajnem, który jest zastosowany w niniejszej predykcji szeregów czasowych z wykorzystaniem MATLAB i rozszerzenia Curve Fitter. W kontekście implementacji w MATLAB, splajn wygładzający  $s$  jest

konstruowany dla określonego parametru wygładzania  $p$  i określonych wag  $w_i$ . Celem jest minimalizacja funkcji:

$$p \sum_i w_i (y_i - s(x_i))^2 + (1 - p) \int \left( \frac{d^2 s}{dx^2} \right)^2 dx \quad (56)$$

Wagi nie są określone, zakłada się, że są równe 1 dla wszystkich punktów danych.

Model porównuje dopasowanie interpolującego splajnu kubicznego i splajnu wygładzającego poprzez optymalizację parametrów  $kwst$  i  $p$ . Miarą dokładności jest SQE. Parametr  $kwst$  oznacza kroki wstecz, analogicznie jak w przypadku regresji wielomianowej.

Parametr wygładzania  $p$  jest zdefiniowany w zakresie od 0 do 1 z inkrementacją co 0,01. Wartość  $p = 0$  daje dopasowanie liniowe metodą najmniejszych kwadratów, a  $p = 1$  daje interpolujący splajn kubiczny. („Smoothing Splines - MATLAB & Simulink”, <https://www.mathworks.com/help/curvefit/smoothing-splines.html>, 2023).

Algorytm predykcji został zaimplementowany w sposób, w którym dla każdej kombinacji wartości  $kwst$  w zakresie liczba całkowitych 2-25 i  $p$  w (zakresie 0,01 do 1 z inkrementacją 0,01) wykonuje następujące operacje:

1. Wybiera dane do trenowania ( $Y_{train}$ ) na podstawie aktualnego  $kwst$ .

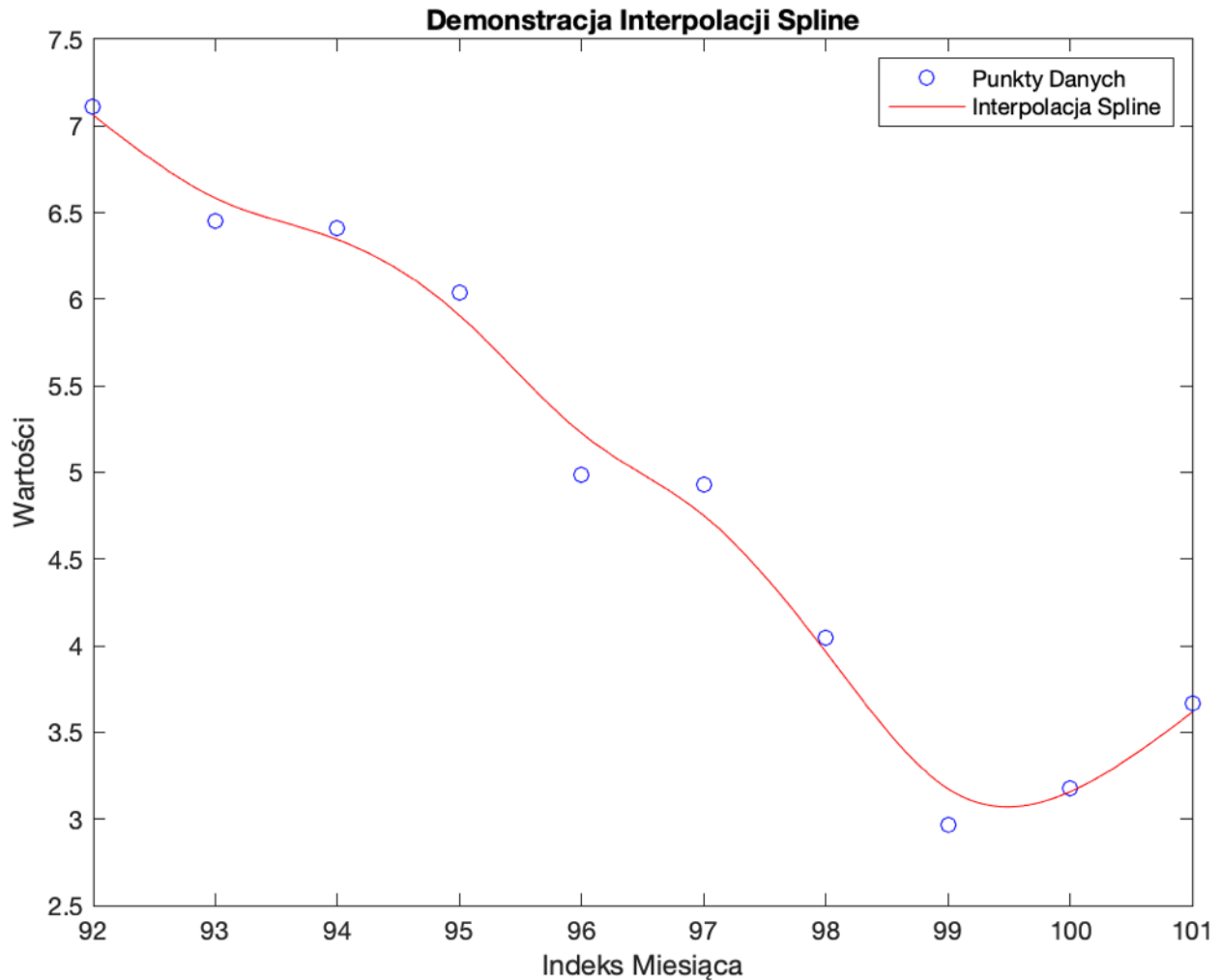
$$Y_{train} = Y[i - kwst + 1:i]. \quad (57)$$

Gdzie  $Y$  to wektor szeregu czasowego a  $i$  to bierzący indeks w procesie iteracji.

```
Y_train = Yn(max(1, i-kwst+1):i);
```

Wykorzystuje funkcję **fit** z MATLAB do dopasowania modelu splajnu wygładzającego splineFit do danych treningowych  $Y_{train}$ , używając aktualnego parametru wygładzania  $p$  i wektora czasu timeVector odpowiadającego wektorowi danych  $Y_{train}$ , co zobrazowano na Rys. 38. .

```
timeVector = (1:length(Y_train))';
% Dopasowanie splajnu z parametrem p
splineFit = fit(timeVector, Y_train, 'smoothingspline', 'SmoothingParam',
p);
```



Rys. 38. *Zobrazowanie interpolacji splajnu. Punkty oznaczone niebieskimi kółkami oznaczają wartości rzeczywiste szeregu czasowego, czerwona linia to przebieg splajnu. Źródło: opracowanie własne.*

2. Prognozowanie następnego punktu odbywa się poprzez dopasowany model splajnu do prognozowania wartości w kolejnym punkcie czasowym.

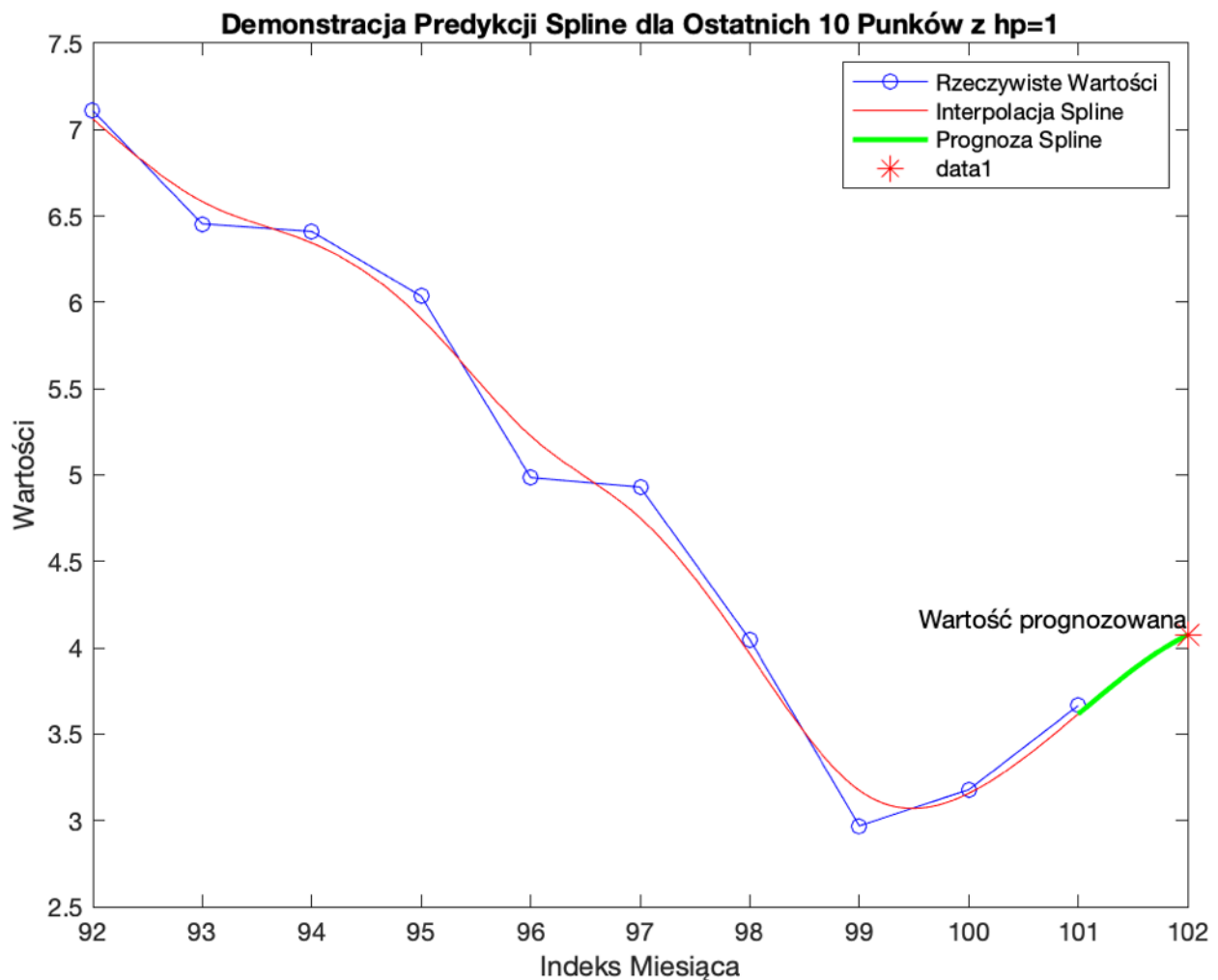
```
% Prognozowanie dla kolejnego kroku
nextTimePoint = length(Y_train) + 1;
splinePrediction = feval(splineFit, nextTimePoint);
```

3. Następnie oblicza się SQE między prognozowaną a rzeczywistą wartością w następnym punkcie czasowym. Jeśli obliczony SQE jest mniejszy niż

dotychczasowy minimalny SQE, aktualizuje optymalne wartości  $kwst$ ,  $p$  i prognozy.

- Następnie, znając optymalne wartości parametrów  $kwst$  i  $p$  wydłuża się wektor czasu o  $hp=1$  i dokonuje predykcji kolejnego elementu, co zobrazowano na Rys. 39.

```
splineFitOptimal = fit(timeVectorOptimal, Y_train_optimal,
'smoothingspline');
Yp_next = feval(splineFitOptimal, length(Y_train_optimal) + hp);
```



Rys. 39. Demonstracja predykcji regresją wygładzającą splajnem. Niebieskie punkty połączone linią to wartości rzeczywiste, czerwona linia to interpolowany splajn, zielona linia to ekstrapolacja splajnu, czerwona gwiazdka oznacza prognozowaną wartość w  $hp=1$ . Źródło: opracowanie własne.

Zaproponowaną predykcję z optymalizacją parametrów  $kwst$  i  $p$  porównano z domyślnymi parametrami funkcji **fit** środowiska MATLAB. Domyślny parametr wygładzania  $p$  dla funkcji 'smoothingspline' w MATLABie jest wybierany w interesującym zakresie, który często można przybliżyć wzorem (57-1):

$$p \approx \frac{1}{\left(1 + \frac{h^3}{6}\right)} \quad (57-1)$$

gdzie:

$h$  - jest średnim odstępem między punktami danych. Jest to metoda heurystyczna, której celem jest zbalansowanie między dokładnością dopasowania a gładkością krzywej.

W kontekście predykcji omawianego modelu wartość  $h$  można obliczyć ze wzoru (57-2).

$$h = \frac{\sum_{i=1}^n (y_{i+1} - y_i)}{n - 1} \quad (57-2)$$

gdzie:

$y_i$  -  $i$ -ta wartość w szeregu danych,

$y_{i+1}$  to wartość następująca po  $y_i$ ,

$n$  - liczba wartości w szeregu danych.

### 3 Badanie dokładności zaprezentowanych metod predykcji rozprzestrzeniania się wirusa.

Rozdział rozpoczyna się opracowaniem planu prac badawczych z wykorzystaniem opisanej adaptacji metody GQM. Dla celu pracy w zakresie pandemii zostają postawione określone pytania mierzalne metryką MAPE, inaczej niż w procentowym formacie, tym razem w formacie dziesiętnym. Do każdego pytania zostaje przypisana metoda i kryteria. Badanie każdą metodą zostało udokumentowane dokładnie co najmniej jednokrotnie dla poszczególnej metody. Szczególnie interesujące w ocenie autora jest udokumentowanie działania metody multiregresji z pseudoinwersją Moore'a-Penrose'a.

Pierwszym badaniem jest wpływ szczepień na dokładność predykcji. Kolejne to badanie dokładności predykcji w zależności od długości horyzontu predykcji oraz liczby sąsiadów-krajów o zbliżonej dynamice zakażeń. Ustalona optymalna wartość liczby sąsiadów została wykorzystana w kolejnych badaniach. W kolejnym kroku zbadano jednocześnie dokładności metod predykcji w

tożsamym interwale czasowym w zależności od długości horyzontu predykcji. Wybór danych był wyzwaniem dla metod, bowiem wybrano ekstremalny przypadek fali zakażeń w Indiach. Badanie zakończono zbiorczym zestawieniem wyników i podsumowaniem. Następnie dokonano badania wpływu położenia geograficznego na dokładność predykcji. Dla wybranych 4 kontynentów wybrano po 4 kraje położone blisko siebie. Wykryto i opisano anomalię w danych z krajów z Afryki. Badanie kończy zestawienie wyników w tabeli oraz ich graficzna interpretacja wraz z podsumowaniem. Ostatnim badaniem predykcji w szeregach czasowych liczby zakażeń było zbadanie jej skuteczności w 4 zdefiniowanych fazach pandemii. Wybór padł na ekstremalny przypadek ze względu na dynamikę rozwoju pandemii na Indie. Zobrazowano wyniki badania oraz przedstawiono w postaci tabeli wraz z podsumowaniem.

### 3.1 Plan prac badawczych

Plan prac badawczych został opracowany z zastosowaniem adaptacji podejścia GQM (ang. Goal Question Metric) (Van Solingen et al. (2002)). Jest to metoda stosowana powszechnie do uzyskania informacji zwrotnej oraz ewaluacji w inżynierii oprogramowania. Jej główną cechą jest zorientowanie na cel, w tym przypadku cel niniejszej rozprawy doktorskiej. Główne elementy a zarazem etapy podejścia do badania to kolejno:

- Cel (Goal) - powinien odpowiadać SMART (określenie kryteriów w celu ustalania celów). SMART oznacza:
  - Specyficzny (Specific)
  - Mierzalny (Measurable)
  - Osiągalny (Achievable)
  - Realistyczny (Realistic)
  - Powiązany z czasem (Time bound)
- Pytanie (Question) - wykorzystanie zbioru pytań do określenia obszaru zainteresowania do analizy. Pytania powinny być powiązane z celami i prowadzić do uzyskania szukanej informacji.
- Metryka (Metric) - metryki są wykorzystywane do mierzenia odpowiedzi na stawiane pytania. Każde pytanie musi być mierzalne za pomocą metryki.

Istotnym elementem techniki GQM są ewaluacja, monitorowanie i doskonalenie tworzonych metryk. Podejście GQM w celu zebrania metryk unika decydowania o przeznaczeniu zebranych już danych i szukania w nich wartości, tzw. łowienia rezultatów. Łączy dane z celami.

Podejście GQM opisano w poniższych krokach:

1. Tworzenie zbioru celów
2. Tworzenie pytań określających cele w sposób mierzalny (ilościowy)
3. Określenie metod badań oraz definicja metryk niezbędnych do odpowiedzi na postawione pytania
4. Utworzenie mechanizmu do zbioru danych
5. Zbieranie, walidacja oraz analiza danych
6. Wykorzystanie danych do dalszych usprawnień

Jedynym celem uwzględnianym w tej metodzie jest cel niniejszej pracy badawczej. Pytania określające cel zostały zdefiniowane następująco:

Q1: Która z metod jest bardziej dokładna w tożsamym zakresie czasu dla badanego kraju?

Q2: Czy uwzględnienie szczepień ma wpływ na predykcję?

Q3: Czy dokładność predykcji dla danego kraju jest podobna do grupy krajów z tego samego regionu geograficznego?

Q4: Czy dokładność predykcji dla danego kraju jest zachowana dla wybranych faz pandemii?

Q5: Jak zmienia się dokładność predykcji w zależności od horyzontu predykcji dla wybranego kraju?

Q6: Jak wpływa liczba najbliższych sąsiadów na dokładność predykcji?

Pomiary zostaną wykonane z wykorzystaniem proponowanych metod badawczych. W paragrafie tym zostanie przedstawione przynajmniej jedno pełne wykorzystanie każdej z 3 metod badawczych, dla których pokazane zostaną istotne dostępne wykresy, tabele danych, wartości macierzy i inne dane obliczeniowe. Dla kolejnych uruchomień metod badawczych wykresy i tabele zostaną ograniczone tylko do elementów z istotnymi obserwacjami z badań. Należy zwrócić uwagę, że metody różnią się od siebie zasadniczo (Tab. 23. ) i ze względu na ich ograniczenia nie wszystkie można zastosować do danego badania.

Tab. 23. Porównanie metod wykorzystanych w badaniu (+ oznacza spełnienie kryterium, - oznacza brak spełnienia kryterium) . Źródło: opracowanie własne.

Kryterium	Metoda badania		
	Multiregresja z pseudoinwersją Moore'a-Penrose'a	Metoda oparta na łańcuchach Markowa	n- najbliższych sąsiadów (z średnią krocząca)
<b>Uwzględnienie szczepień</b>	+	-	-
<b>Niezależność od danych z innych krajów</b>	-	+	-
<b>Ograniczenia danych wejściowych</b>	-	+	+
<b>Zmienne niezależne</b>	wiele (wymagane informacje o szczepieniach)	jedna	wiele
<b>Predykcja liczby zakażeń</b>	dobowe skumulowane	dobowe przyrosty	dobowe skumulowane
<b>Zalecana długość okna uczenia</b>	Zmienna, zależna od horyzontu predykcji i szerokości macierzy uczenia	tygodniowe okna czasowe, zalecana ilość okien >4	zalecana 3x wielokrotność okna predykcji

Dla każdego ze zdefiniowanych pytań przypisano metodę badawczą (Tab. 24. )

Tab. 24. Przepisanie metod badawczych do określonego (za pomocą podejścia GQM) pytania.

Kryterium	Metoda badania		
	Multiregresja z pseudoinwersją Moore'a- Penrose'a	Metoda oparta na łańcuchach Markowa	n- najbliższych sąsiadów (z średnią krocząca)
<b>Q1: Która z metod jest bardziej dokładna w tożsamym zakresie czasu i horyzoncie predykcji dla badanego kraju?</b>	•	•	•
<b>Q2: Czy uwzględnienie szczepień ma wpływ na predykcję?</b>	•		
<b>Q3: Czy dokładność predykcji dla danego kraju jest podobna do grupy krajów z tego samego regionu geograficznego?</b>	•	•	•
<b>Q4: Czy dokładność predykcji dla danego kraju jest zachowana dla wybranych faz pandemii?</b>	•		•
<b>Q5: Jak zmienia się dokładność predykcji w zależności od horyzontu predykcji dla wybranego kraju z zastosowaniem danej metody?</b>	•	•	•
<b>Q6: Jak wpływa liczba najbliższych sąsiadów na dokładność predykcji?</b>			•

Metryką stosowaną do badania i porównania dokładności predykcji stosowanych metod jest MAPE. W badaniach zastosowano pewne ułatwienie, zgodne z podejściem GQM, mające na celu obliczenia oraz pokazanie wartości dokładności predykcji. Mianowicie MAPE i APE, pomimo

bycia jednostką wyrażaną w procentach, tutaj jest przedstawione w formie dziesiętnej. W celu otrzymania wartości procentowej, każdą przedstawioną wartość w postaci dziesiętnej należy zastosować iloczyn wartości dziesiętnej i 100%. Przykładowo MAPE w formie dziesiętnej 0,1234 odpowiada 12,34 %.

Zbieranie danych odbywać się będzie za pomocą autorskiego rozwiązania opracowanego w języku programowania Python. Dane dostępne online pobierane są z repozytorium w sposób opisany w paragrafie 2.3.1. Zależnie od wybranej metody badawczej oraz parametrów wejściowych algorytmu ujętych w Tab. 25. dane są walidowane pod względem ograniczeń i założeń danej metody. W przypadku braku spełnienia ograniczeń badanie metodą jest przerywane lub algorytm dostosowuje dane wejściowe w opisanym zakresie dla danej metody.

*Tab. 25. Dane i parametry algorytmów dla wybranych badań. (Źródło: mpracowanie własne).*

<b>Badanie</b>	<b>Dane i parametry algorytmu</b>
<b>Q1: Która z metod jest bardziej dokładna w tożsamym zakresie czasu i horyzoncie predykcji dla badanego kraju?</b>	Wybrany jeden kraj Horyzont predykcji: 1, 3, 7, 14, 21 dni
<b>Q2: Czy uwzględnienie szczepień ma wpływ na predykcję?</b>	Porównanie dokładności predykcji dla jednego z krajów o niskiej i wysokiej gęstości szczepień
<b>Q3: Czy dokładność predykcji dla danego kraju jest podobna do grupy krajów z tego samego regionu geograficznego?</b>	Predykcja na podstawie wybranego okresu czasu uczenia i horyzontu predykcji dla czterech sąsiadujących krajów na czterech wybranych kontynentach.
<b>Q4: Czy dokładność predykcji dla danego kraju jest zachowana dla wybranych faz pandemii?</b>	Predykcja w horyzoncie predykcji dla czterech faz pandemii w czterech krajach.
<b>Q5: Jak zmienia się dokładność predykcji w zależności od horyzontu predykcji dla wybranego kraju z zastosowaniem danej metody?</b>	Na podstawie wyników badania Q1
<b>Q6: Jak wpływa liczba najbliższych sąsiadów na dokładność predykcji?</b>	Dla parametrów n: 3, 5, 9, 13 i horyzontu predykcji: 3, 7, 14, 21

Kolejnym krokiem jest uruchomienie zaimplementowanego w języku Python algorytmu danej metody oraz zapisanie wyników i obserwacji. Zastosowanie skryptów do danego badania oraz ewentualne modyfikacje będą zgodne z ich przypisaniem określonym w Tab. 24. Modyfikacje nie będą dotyczyć działania samych algorytmów, ale operacji na danych oraz parametrów uruchomieniowych (generowanie wykresów, długość okna uczenia, zakres dat horyzontu predykcji, operacje na danych wejściowych, weryfikacja poprawności danych, itp.). Modyfikacje nie będą dotyczyły algorytmów.

### 3.2 Badanie wpływu gęstości szczepień na dokładność predykcji.

Badanie dokładności predykcji z uwzględnieniem szczepień oparto na zaproponowanym wskaźniku gęstości szczepień i stanowi ono realizację zadań związanych z pytaniem Q2 określonym w planie prac badawczych w podrozdziale 3.1. Założeniem było przeprowadzenie badania dla krajów o niskiej i wysokiej wartości tego wskaźnika oraz porównanie metryk MAPE. Istotne jest wykonanie badania w identycznym zakresie czasowym pod względem okna uczenia oraz horyzontu predykcji.

Na wstępie należy wybrać zakres dat, dla którego będą dostępne dane o szczepieniach zgodnie z kryteriami opisanymi w metodzie. Przyjęto zakres dat okna uczenia od 1 do 21 sierpnia 2021 roku. Okres predykcji to id 21 do 28 sierpnia.

Kraje wybrano na podstawie tabeli gęstości szczepień dla pierwszego dnia okna uczenia (Tab. 26. ), czyli stan na 1 sierpień 2021 roku. W celu wykorzystania w pełni możliwości metody, wybrano Islandię jako 4 kraj od początku listy – z największą gęstością szczepień. Krajem o mniejszej gęstości szczepień znajdującym na 4 pozycji od końca listy gęstości szczepień okazała się Uganda.

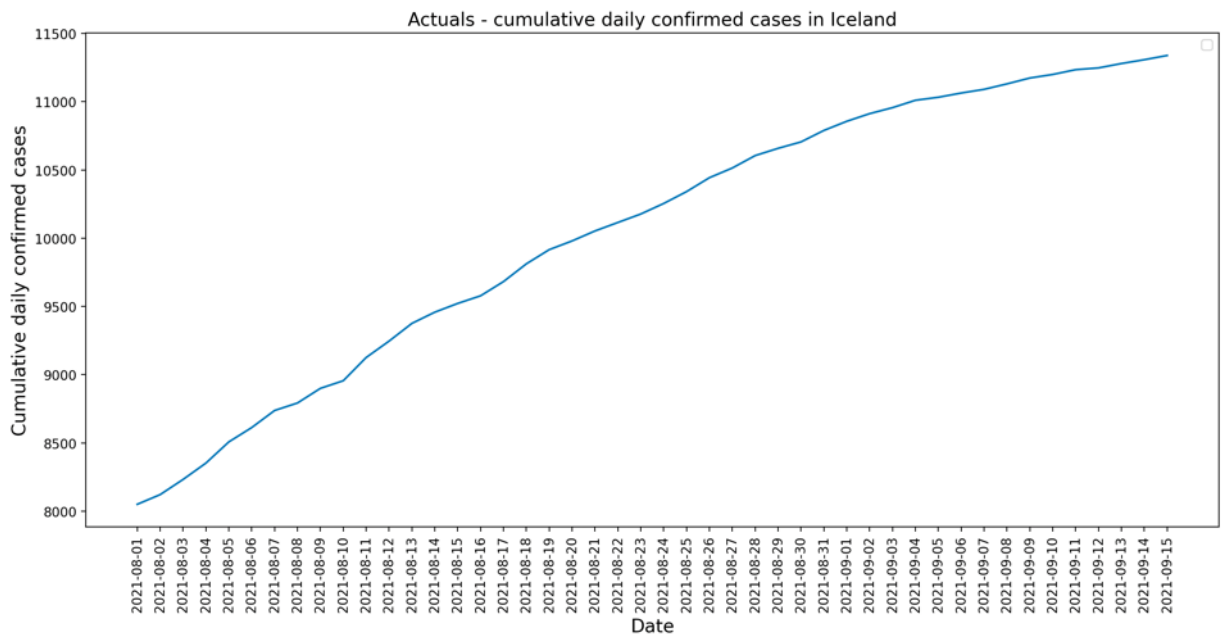
Tab. 26. Lista krajów o największej i najmniejszej gęstości szczepień na dzień 1 sierpnia 2021 r. Kolorem zielonym oznaczono kraj o wysokiej wartości gęstości szczepień - Islandię, natomiast kolorem czerwonym kraj o niskiej wartości gęstości szczepień - Ugandę. Źródło: opracowanie własne na podstawie CSSE JHU.

Kraj	Populacja	1.08.2021	Gęstość szczepień
Malta	441539	759818	1720,8
United Arab Emirates	9890400	16810996	1699,7
Seychelles	98340	141435	1438,2
Iceland	341250	470374	1378,4
Uruguay	3473727	4788109	1378,37804
Bahrain	1701583	2314773	1360,4
San Marino	33938	45583	1343,1
Chile	19116209	25642416	1341,4
...			
Kenya	53771300	1726193	32,1
Sierra Leone	7976985	225380	28,3
Malawi	19129955	504195	26,4
Uganda	45741000	1143763	25,0
Mozambique	31255435	691040	22,1
Nigeria	206139587	3938945	19,1
Gambia	2416664	43557	18,0

#### **Badanie dokładności predykcji dla kraju o wysokiej gęstości szczepień.**

Zgodnie z tabelą gęstości szczepień dla początku okna uczenia (1 sierpień 2021 r.) wybrano kraje: Malta, Zjednoczone Emiraty Arabskie, Seszele, Urugwaj, Bahrajn i San Marino.

Przyjęty horyzont predykcji określono na poziomie 7 dni. Następnym krokiem była wizualizacja przebiegów rzeczywistych szeregu czasowego skumulowanych dobowych zakażeń (Rys. 40. ).



Rys. 40. Wykres przebiegu czasowego skumulowanej dobowej liczby zakażeń w Islandii od 1 sierpnia do 15 września 2021 r. Źródło opracowanie własne.

Pseudoinwersja Moore’a – Penrose’a jest na tyle kluczowym i interesującym elementem tej metody, że autor niniejszej rozprawy zdecydował się przedstawić każdy krok i przekształcenie macierzy metody zastosowanej w prognozowaniu rozprzestrzeniania się wirusa.

Równanie opisujące model (29) uwzględnia wektor  $\vec{y}$  (Tab. 28.) i macierz C przedstawiona w Tab. 27. Poszukiwany jest wektor  $\vec{A}$ . Dla przypomnienia, równanie to przyjmuje postać:

$$\vec{y} = C\vec{A} \quad (29)$$

Tab. 27. Macierz  $C$  z konkatenowanym wektorem  $\mathbf{1}$  (zgodnie z opisem w Tab. 11. ) dla

		United Arab Emirates					San Marino
	Ones	Malta	Emirates	Seychelles	Uruguay	Bahrain	Marino
1.08.2021	1	34375	682377	18362	381569	269303	5143
2.08.2021	1	34439	683914	18411	381715	269401	5147
3.08.2021	1	34508	685462	18470	381853	269495	5167
4.08.2021	1	34590	686981	18546	381994	269617	5178
5.08.2021	1	34668	688489	18617	382155	269737	5179
6.08.2021	1	34716	690009	18617	382295	269848	5183
7.08.2021	1	34787	691554	18617	382360	269949	5183
8.08.2021	1	34846	692964	18617	382506	270060	5183
9.08.2021	1	34902	694285	18617	382607	270161	5194
10.08.2021	1	34953	695619	18668	382721	270290	5207
11.08.2021	1	35040	696906	18772	382873	270451	5210
12.08.2021	1	35091	698166	18954	382997	270590	5217
13.08.2021	1	35192	699381	18981	383101	270692	5222
14.08.2021	1	35286	700587	19082	383212	270816	5222

badanego przypadku. Źródło: opracowanie własne.

Tab. 28. Wektor  $\vec{y}$  reprezentujący państwo będące zmienną objaśnianą  $Y$ , który należy przesunąć o horyzont prognozy  $h_p = 7$  w celu zbudowania modelu i umożliwienia predykcji zgodnie z opisem w 0 Źródło: opracowanie własne.

	$\vec{y}$
12.09.2021	11247
13.09.2021	11279
14.09.2021	11307
15.09.2021	11338
16.09.2021	11377
17.09.2021	11404
18.09.2021	11404
19.09.2021	11404
20.09.2021	11471
21.09.2021	11518
22.09.2021	11558
23.09.2021	11594
24.09.2021	11632
25.09.2021	11632

Wektor  $\vec{A}$  obliczany jest zgodnie z (30) z wykorzystaniem macierzy  $C^+$  będącej wynikiem pseudoinwersji Moore'a-Penrose'a macierzy  $C$ . Macierz  $C^+$  w oryginalnej postaci dla omawianego przykładu została przedstawiona w Tab. 29.

Tab. 29. Macierz  $C^+$  powstała poprzez pseudoinwersję Moore'a-Penrose'a macierzy  $C$  zgodnie z (30). Źródło: opracowanie własne.

-512,815	-708,276	990,903	1453,7	-1009,89	-3207,11	4469,14	-170,056	277,472	783,241	-3443,54	607,02	-234,814	706,012
0,0008337	-0,0025309	3,64E-06	0,00314042	0,0030136	-0,0014936	-0,0019428	-0,0006423	0,0028985	-0,0010484	0,00279011	-0,01677	0,00272883	0,00901915
-0,0002516	-7,42E-05	0,00017932	0,00014918	-0,0003744	-0,0007232	0,00132114	0,00014873	4,45E-05	0,00021627	-0,0010363	0,00078797	-0,0002148	-0,0001726
-0,0018393	0,00022665	0,00059067	0,00051336	0,00033142	-0,0009304	0,00489195	0,0006243	-0,0018139	-0,0018932	-0,0056395	0,00539821	-0,0006582	0,00019784
-0,0016821	0,00120833	-0,0030913	-0,0039073	0,0052503	0,0112158	-0,0103566	0,00292194	-0,0001613	-0,0039258	0,00729195	-0,002624	-0,0003704	-0,0017695
0,00500684	0,00181889	-0,0001139	-0,001232	-0,0031989	-0,0017654	-0,0051416	-0,0032844	-0,0011217	0,0019488	0,00487516	0,00144037	0,0014873	-0,0007196
-0,0031696	-0,0211407	0,0167462	0,0291363	0,00267467	-0,0068569	-0,011147	-0,0294146	-0,0020543	0,0219985	0,0122793	-0,0105176	0,00787312	-0,0064075

Postać wektora po przekształceniu  $C^+ \vec{y} \approx \vec{A}$  (30) przyjmuje postać przedstawioną w Tab. 30.

Tab. 30. Obliczony wektor  $\vec{A}$  współczynników równania regresyjnego pozwalający na wyznaczenie modelu  $Y_m$  zgodnie z (30). Źródło: opracowanie własne

-184002
-0,106898
-0,0389204
-0,124817
0,207198
0,511791
2,11765

Tab. 31. Macierz  $C_y$  powstała w analogiczny sposób jak macierz  $C$  z tą różnicą, że indeks wierszy szeregu czasowego jest przesunięty w czasie w przód o liczbę wierszy odpowiadający wartości horyzontu predykcji  $hp$ , w tym przypadku  $hp = 7$ . Źródło: opracowanie własne.

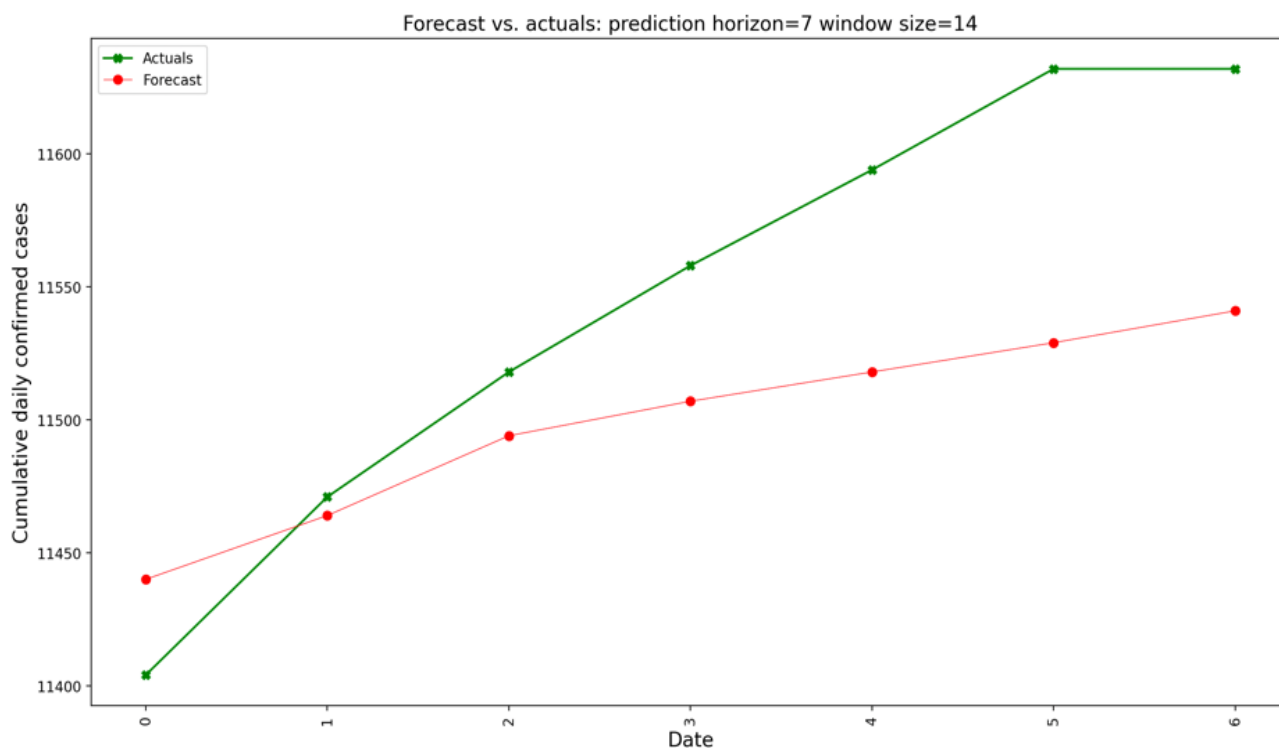
	Ones	United		Seychelles	Uruguay	Bahrain	San
		Malta	Emirates				Marino
8.08.2021	1	34846	692964	18617	382506	270060	5183
9.08.2021	1	34902	694285	18617	382607	270161	5194
10.08.2021	1	34953	695619	18668	382721	270290	5207
11.08.2021	1	35040	696906	18772	382873	270451	5210
12.08.2021	1	35091	698166	18954	382997	270590	5217
13.08.2021	1	35192	699381	18981	383101	270692	5222
14.08.2021	1	35286	700587	19082	383212	270816	5222
15.08.2021	1	35337	701776	19149	383292	270919	5222
16.08.2021	1	35380	702885	19199	383387	271025	5231
17.08.2021	1	35434	704000	19224	383505	271130	5238
18.08.2021	1	35516	705089	19272	383645	271257	5244
19.08.2021	1	35602	706166	19320	383758	271353	5250
20.08.2021	1	35653	707236	19476	383903	271432	5255
21.08.2021	1	35721	708302	19530	384026	271534	5255

Obliczony wektor  $\vec{A}$  współczynników równania regresyjnego pozwala na wyznaczenie modelu  $Y_m$  zgodnie z równaniem regresyjnym (31). Tym razem wykorzystuje się macierz  $C_y$  (Tab. 31.) powstała w analogiczny sposób jak macierz  $C$  z tą różnicą, że indeks wierszy szeregu czasowego jest przesunięty w czasie w przód o liczbę wierszy odpowiadający wartości horyzontu predykcji  $hp$ , w tym przypadku  $hp = 7$ . Dla przypomnienia, wartości prognozowane stanowią elementy drugiej połowy wektora  $Y_m$  (Tab. 32.).

Tab. 32. Wektor  $Y_m$  (właściwie druga połowa wektora wynikowego zgodnie z Tab. 13. ) stanowiący prognozowane wartości szeregu czasowego w horyzoncie predykcji i będący wynikiem równania regresyjnego (31). Źródło: opracowanie własne.

	$Y_m$
22.08.2021	11404
23.08.2021	11471
24.08.2021	11518
25.08.2021	11558
26.08.2021	11594
27.08.2021	11632
28.08.2021	11632

Następnie obliczany jest błąd APE zgodnie z równaniem (32) – wykorzystując wartości rzeczywiste i prognozowane, których przebieg przedstawiono na Rys. 41. Na podstawie APE obliczane jest MAPE zgodnie z równaniem (33). Widoczne „spłaszczenie” przebiegu danych rzeczywistych pomiędzy przedostatnim a ostatnim dniem świadczy o niskim przyroście dziennym zakażeń w ostatnim dniu.

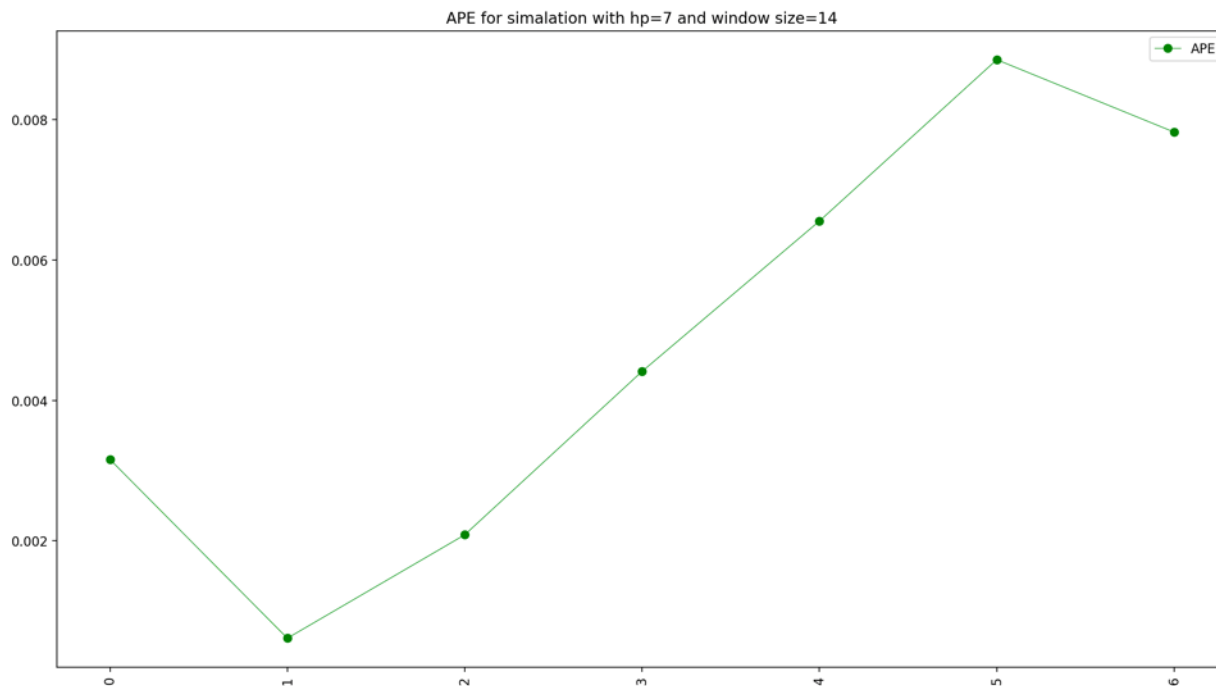


Rys. 41. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji hp: rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Islandii. Oś rzędnych zawiera kolejne dni horyzontu predykcji. Źródło: opracowanie własne.

Wartości APE i MAPE dla kolejnych dni w horyzoncie predykcji przedstawiono w Tab. 33. Przebieg APE przedstawiono na wykresie na Rys. 42.

Tab. 33. Wartości APE i MAPE uzyskane w badaniu dla kraju - Islandia. Źródło: opracowanie własne.

Numer kolejnego dnia w horyzoncie predykcji hp	APE	MAPE
0	0,0031	0,0031
1	0,00061	0,00061
2	0,002	0,0013
3	0,0044	0,0024
4	0,0065	0,0034
5	0,0088	0,0045
6	0,0078	0,005

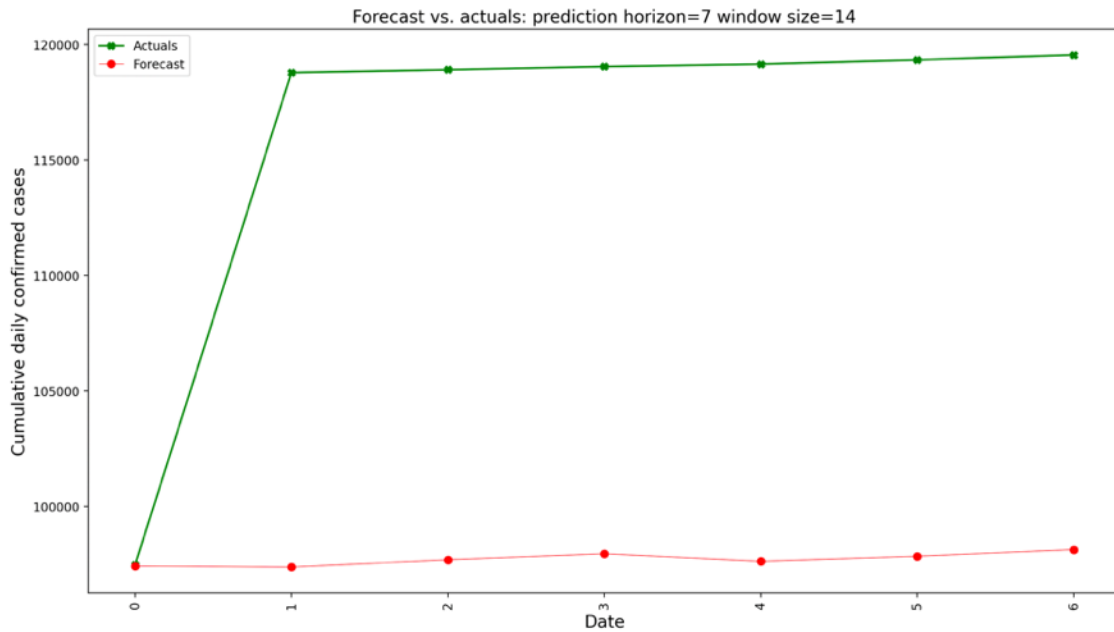


Rys. 42. Przebieg APE dla badanego przypadku - Islandii. Źródło: opracowanie własne.

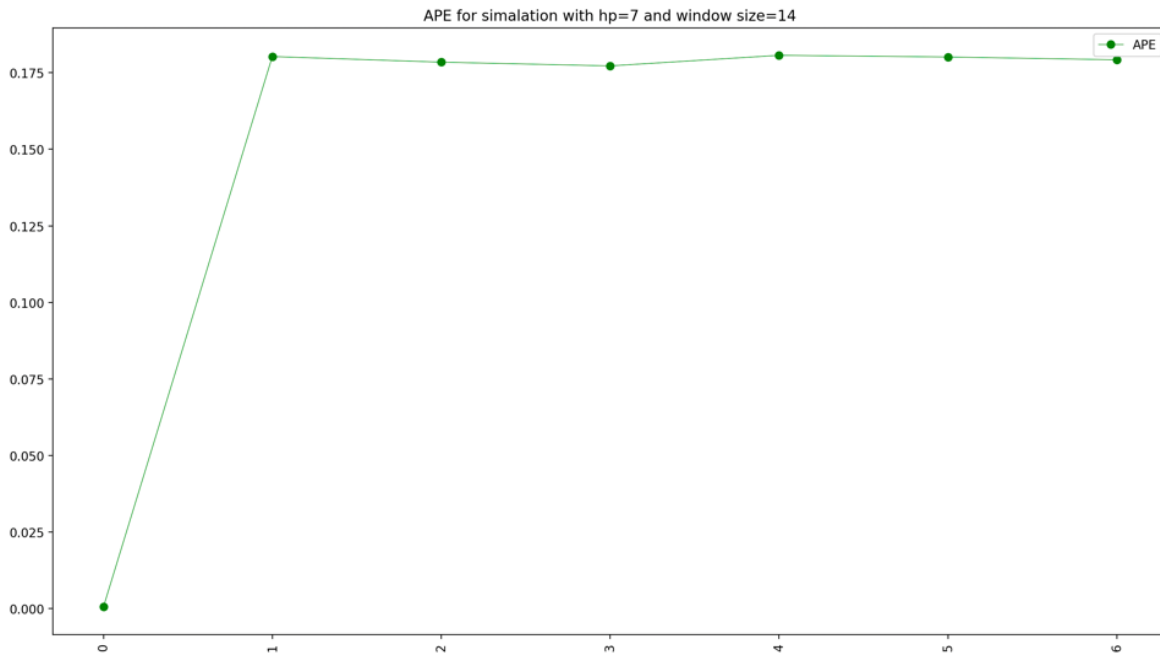
### **Badanie dokładności predykcji dla kraju o niskiej gęstości szczepień.**

Badanie dokładności predykcji dla kraju o niskiej gęstości szczepień wykonano taką samą metodą jak dla kraju o wysokiej gęstości szczepień. Krajami o podobnej gęstości szczepień do referencyjnej Ugandy okazały się kraje leżące w Afryce: Kenia, Sierra Leone, Malawi, Uganda, Mozambik, Nigeria i Gambia.

Prognoza skumulowanej liczby zakażeń w pierwszym dniu horyzontu predykcji wypadła zaskakująco dokładnie osiągając wartość  $APE(0)=0,00048$  (w porównaniu do Islandii  $APE(0)=0,0031$ ). Zobrazowanie wartości rzeczywistych i prognozowanych szeregów czasowych w Rys. 43. i przebieg APE w Rys. 44. pozwalają stwierdzić, że zaskakująco precyzyjna prognoza dla pierwszego dnia w horyzoncie predykcji była prawdopodobnie przypadkowa. Prognozy dla kolejnych dni mają zbliżone wartości APE wynoszące  $\sim 0,18$  (Tab. 34. ).



Rys. 43. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji  $h_p$ : rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Ugandy. Oś rzędnych zawiera kolejne dni horyzontu predykcji. Źródło: opracowanie własne.



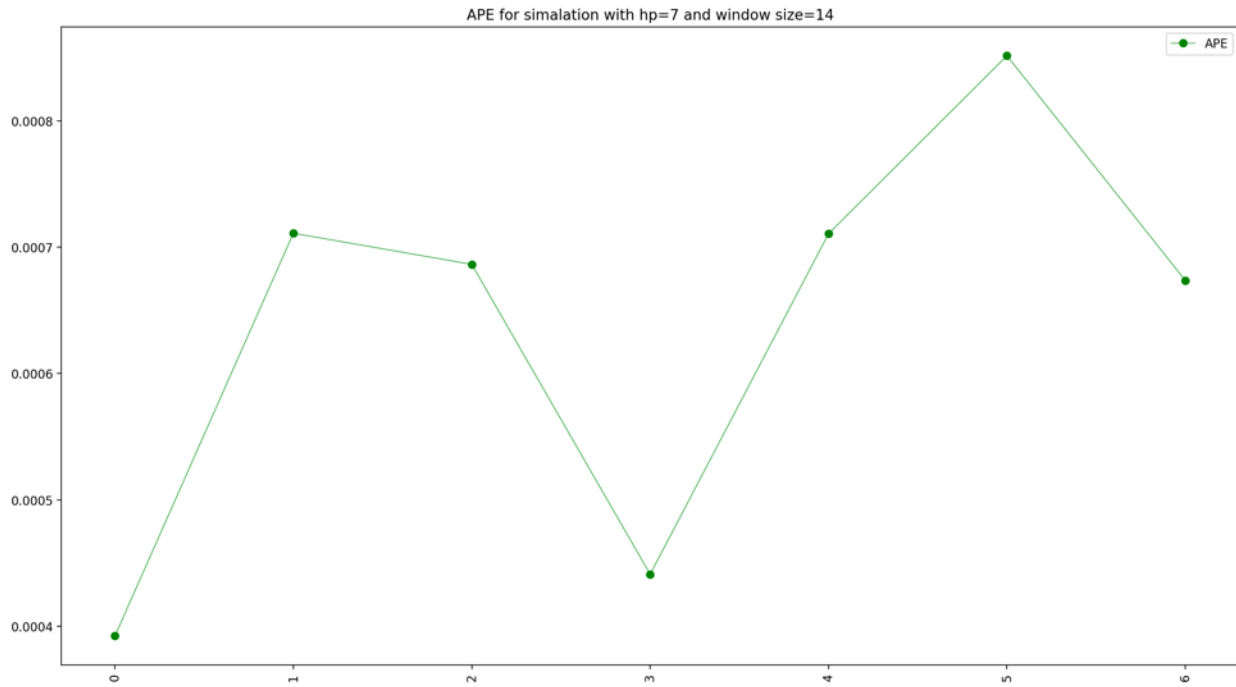
Rys. 44. Przebieg APE dla badanego przypadku - Ugandy. Źródło: opracowanie własne.

Tab. 34. Wartości APE i MAPE uzyskane w badaniu dla kraju - Uganda. Źródło: opracowanie własne.

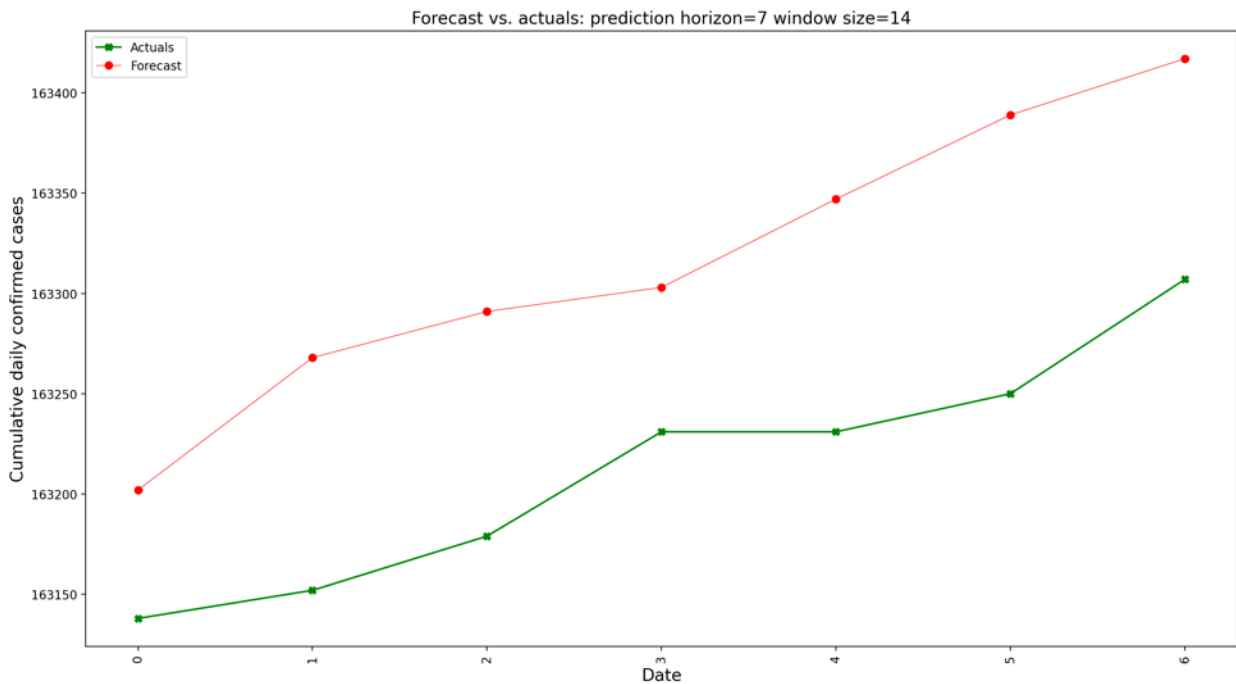
Numer kolejnego dnia w horyzoncie predykcji hp	APE	MAPE
0	0,00048	0,00048
1	0,18	0,18
2	0,178	0,179
3	0,177	0,178
4	0,181	0,179
5	0,18	0,179
6	0,179	0,179

MAPE dla kraju o niskiej gęstości szczepień (Ugandy) w przeprowadzonym badaniu wyniosło 0,18, co jest ~35,5 razy większe niż dla kraju o wysokiej gęstości szczepień (Islandii) wynoszące 0,005.

Badanie powtórzono dla przedziału czasowego od 1 do 28 lutego 2022 r. W badanym okresie dostępność do szczepionek w skali światowej zwiększyła się. Rozważanymi krajami do badania były ponownie Uganda (282) i Islandia (2290) - w nawiasach podano gęstość szczepień na dzień 1 lutego 2022 r. Oba kraje utrzymały się w poprzednich pozycjach tabeli gęstości szczepień.



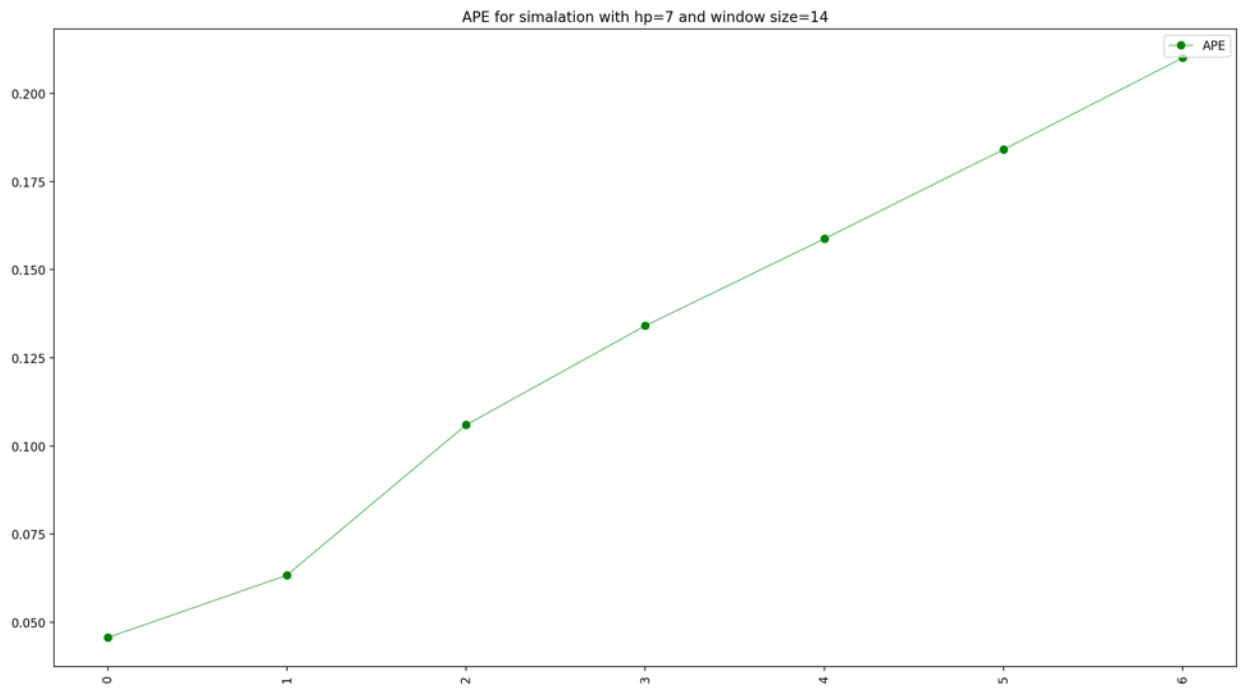
Rys. 45. Przebieg APE dla badanego przypadku – Ugandy, dla zakresu dat: 22-28 luty 2022 r. Źródło: opracowanie własne.



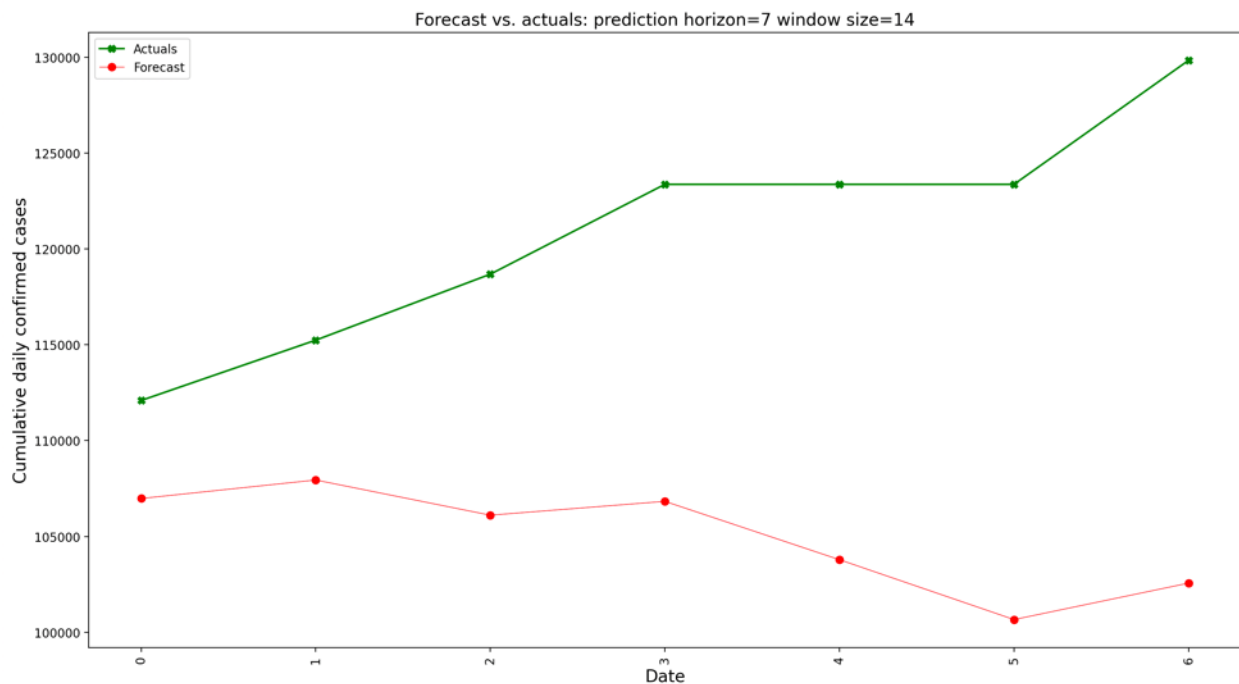
Rys. 46. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji hp: rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Ugandy dla zakresu dat: 22-28 luty 2022 r. Oś rzędnych zawiera kolejne dni horyzontu predykcji. Źródło: opracowanie własne.

Otrzymano dla Ugandy wynik MAPE = 0,00064. Przebiegi APE przedstawiono na wykresie na Rys. 46. Wykonano symulację z takimi samymi zakresami dat i horyzontem predykcji dla Islandii, dla której MAPE = 0,13, co jest ~202 razy większe niż dla Ugandy.

Kolejnym krokiem było porównanie przebiegów rzeczywistych i prognozowanych szeregów czasowych dla obu krajów (Rys. 46. i Rys. 48. ). W przypadku Islandii – prognoza ma przeciwny trend do wartości rzeczywistych, natomiast w przypadku Ugandy – oba przebiegi są niemal równoległe.



Rys. 47. Przebieg APE dla badanego przypadku – Islandii, dla zakresu dat: 22-28 luty 2022 r. Źródło: opracowanie własne.



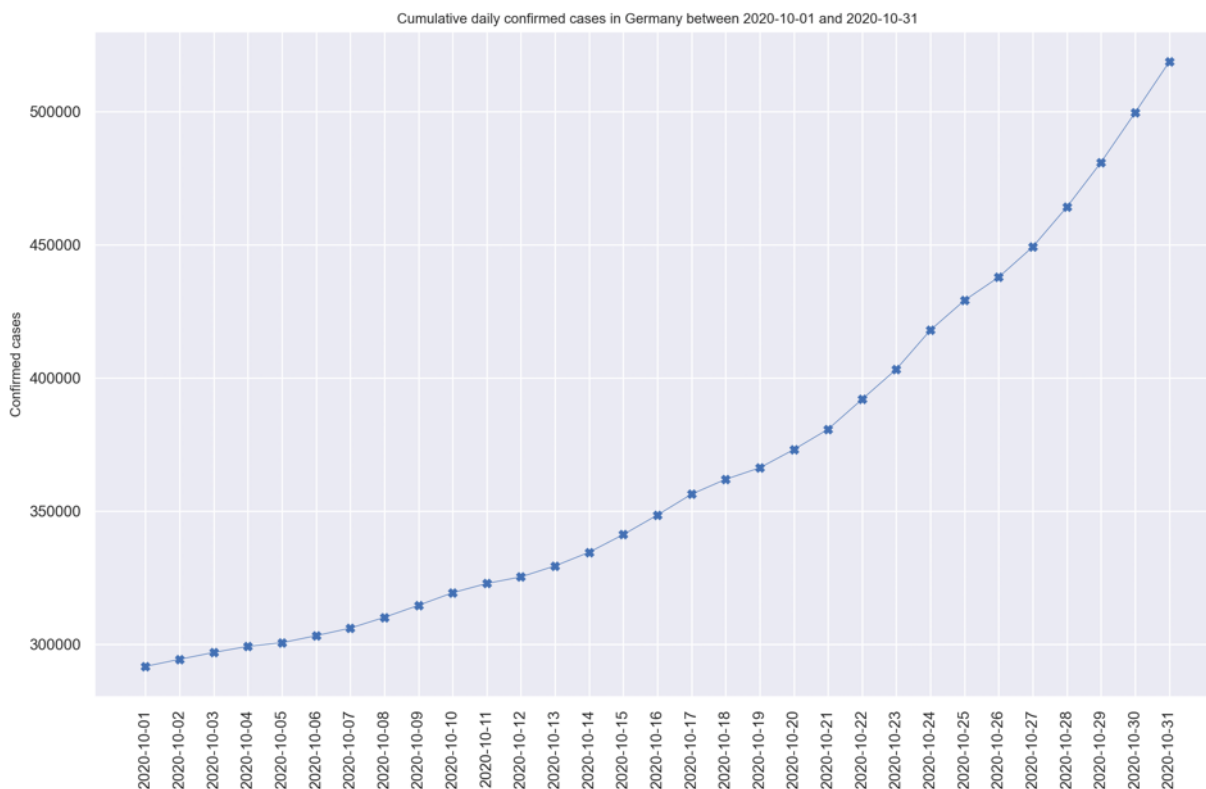
Rys. 48. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji  $h_p$ : rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Islandii dla zakresu dat: 22-28 luty 2022 r. Oś rzędnych zawiera kolejne dni horyzontu predykcji. Źródło: opracowanie własne.

Symulacje dla różnych zakresów dat i krajów utrzymujących swoje pozycje w tabeli gęstości szczepień wykazały przeciwstawne wnioski. Stwierdzić należy zatem, że gęstość szczepień nie ma wpływu na precyzję prognozy.

### 3.3 Badanie wpływu długości horyzontu predykcji i liczby sąsiadów dla określonego kraju.

Badanie wpływu długości horyzontu predykcji i liczby sąsiadów dla określonego kraju stanowi realizację zadań związanych z pytaniem Q6 określonym w planie prac badawczych w podrozdziale 3.1. Metodą wykorzystaną do badania jest metoda oparta o  $n$ -najbliższych sąsiadów i średnią kroczącą. Wynik badania, czyli liczba sąsiadów dająca najdokładniejsze wyniki predykcji będzie wykorzystywana w pozostałych badaniach. Do tego badania wybrano kraj Niemcy. W celu wyboru zakresu dat okresu uczenia wybrano przedział pomiędzy 1 a 30 października 2020 roku.

Przebieg rzeczywistej skumulowanej liczby zakażeń obejmującej badany okres przedstawiono na Rys. 49.



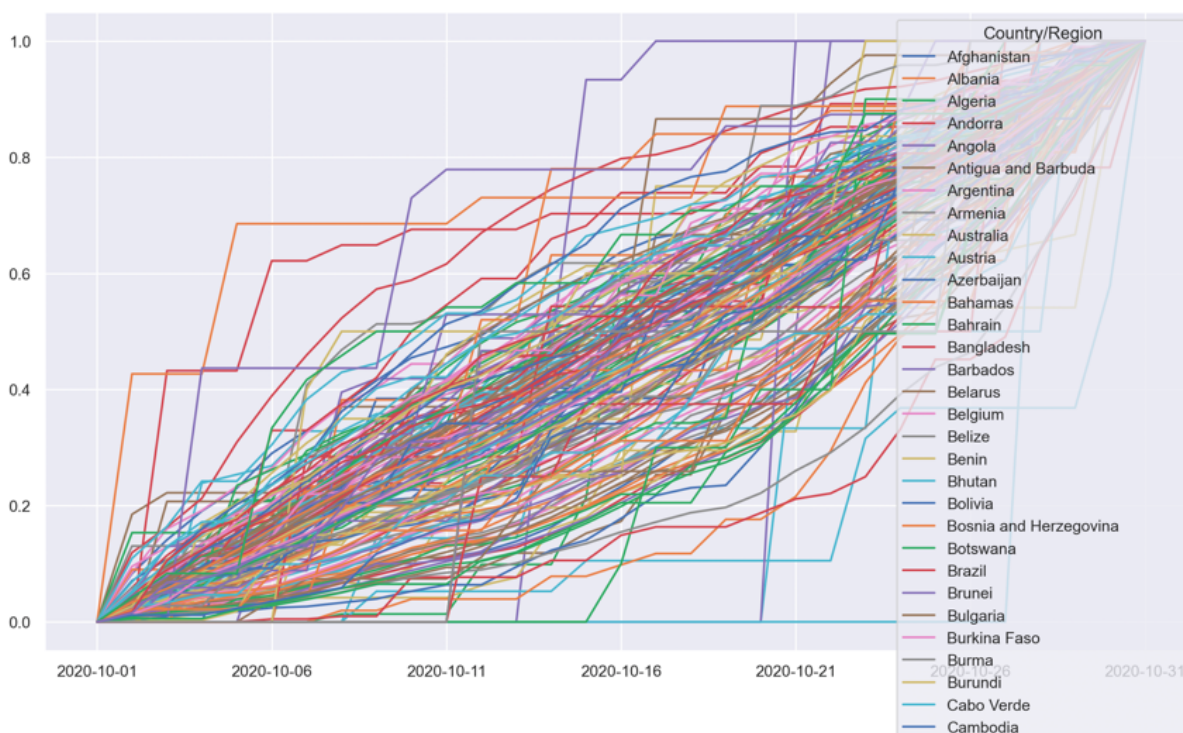
Rys. 49. Skumulowana liczba zakażeń w Niemczech obejmująca badany okres. Źródło: opracowanie własne.

Badanie wykonano dla permutacji parametrów  $n$ -najbliższych sąsiadów ze zbioru wartości (3, 5, 7, 9, 13) oraz horyzontu predykcji  $h_p$  ze zbioru wartości (3, 7, 14, 21).

Poniżej przedstawiono szczegółowo wykonanie badania dla parametru  $n$ -najbliższych sąsiadów  $n=5$  i rozważanych horyzontów predykcji. Przebieg i rezultaty kolejnych badań ograniczono do przedstawienia wybranych wyników, kończąc podsumowaniem uwzględniającym MAPE dla wszystkich kombinacji parametrów.

Pierwszym etapem jest zobrazowanie znormalizowanych rzeczywistych przebiegów kumulatywnej liczby zakażeń dla wszystkich krajów. Cechą wspólną przebiegów jest rozpoczynanie przebiegu z wartości odciętych wynoszących 0, a kończenie na wartości odciętych równej 1. Oś rzędnych stanowi zakres dat z badanego przedziału czasowego (Rys. 50. ). Przebiegi

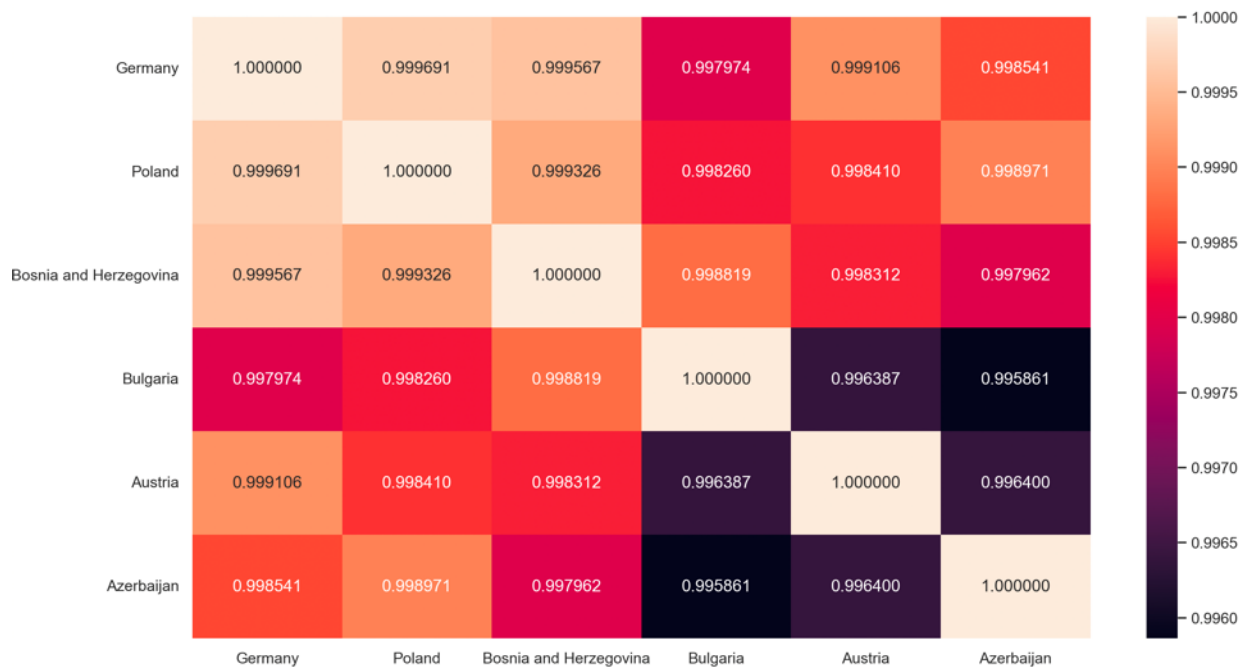
te tworzą malownicze „wstęgi”, których gęstość zwiększa się wraz z bliskością do „przekątnej” wykresu.



Rys. 50. Przebiegi znormalizowanych szeregów czasowych dobowych skumulowanych zakażeń dla wszystkich krajów z dostępnymi danymi. Źródło: opracowanie własne.

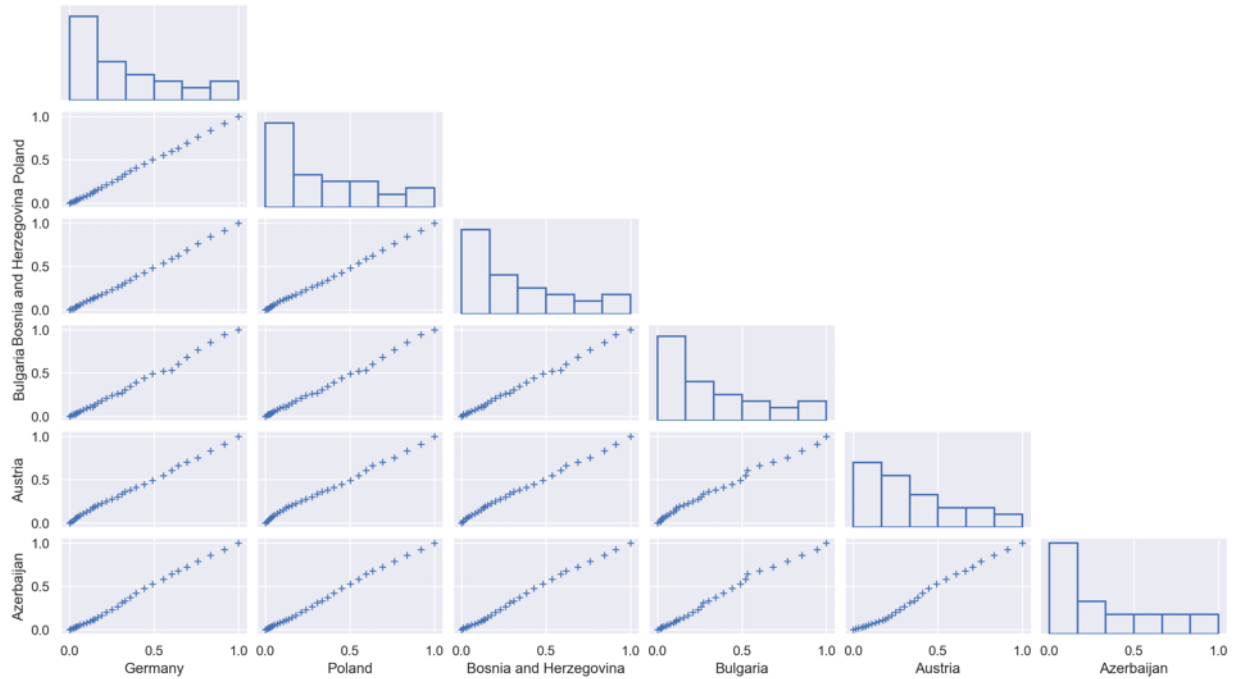
Kolejnym krokiem jest odnalezienie n-najbliższych przebiegów do przebiegu kraju referencyjnego za pomocą metody opisanej w paragrafie 2.3. Selekcja danych na tym etapie jest kluczowa, dlatego skrupulatnie należy sprawdzić poprawność wyboru pod względem podobieństwa do referencyjnego kraju. W badanym przypadku krajami najbliższymi dla referencyjnego przebiegu Niemiec są przebiegi krajów: Polski, Bośni i Hercegowiny, Bułgarii, Austrii, Azerbejdżanu.

Pierwszym krokiem jest zobrazowanie podobieństwa (korelacji wzajemnej) wymienionych krajów za pomocą mapy ciepła Rys. 51. Wartości powinny dążyć do 1. Wartości na przekątnych to korelacja wzajemna tego samego kraju i dlatego wynosi 1.



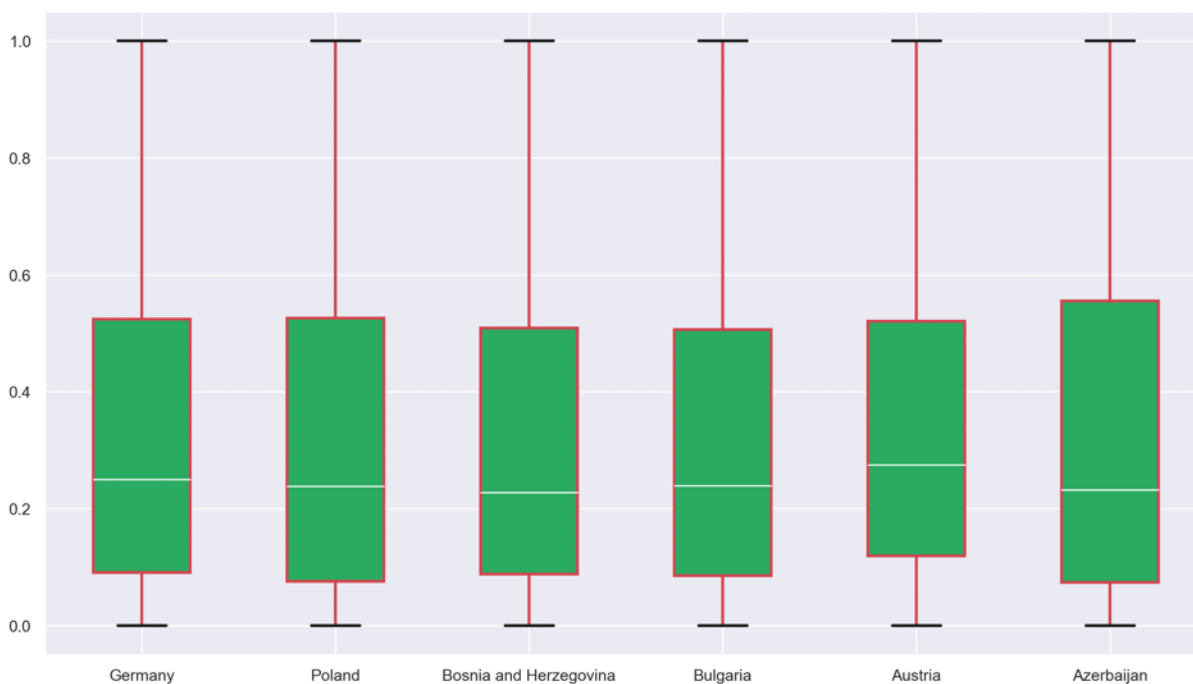
Rys. 51. Mapa ciepła obrazująca korelację wzajemną między krajem bazowym oraz  $n$ -najbliższymi sąsiadami. Źródło: opracowanie własne.

Dopełnieniem analizy opartej na liczbowych wartościach korelacji wzajemnej i korespondującej skali opartej na gradiencie kolorów jest analiza graficzna dwuwymiarowa oraz histogramy (Rys. 52. ). Jeśli przebiegi szeregów czasowych są zbieżne, to dwuwymiarowe zobrazowanie na wykresie w postaci punktów powinno być bliskie prostej biegnącej po przekątnej wykresu.



Rys. 52. Wykres rozkładów dwuwymiarowych par w zbiorze danych reprezentujących kraj bazowy i  $n$ -najbliższych sąsiadów. Źródło: opracowanie własne.

Kolejnym krokiem jest sprawdzenie obecności wartości ponadnormatywnych, tzw. „outlierów.” Wykonuje się to za pomocą wykresu pudełkowego (Rys. 53.) kraju referencyjnego i jego  $n$ -najbliższych sąsiadów. Wykres ten również umożliwi porównanie wartości minimalnych, maksymalnych, mediany i percentylów. Oczekiwany rezultat jest brak „outlierów” oraz podobieństwo wykresu dla zobrazowanych krajów.

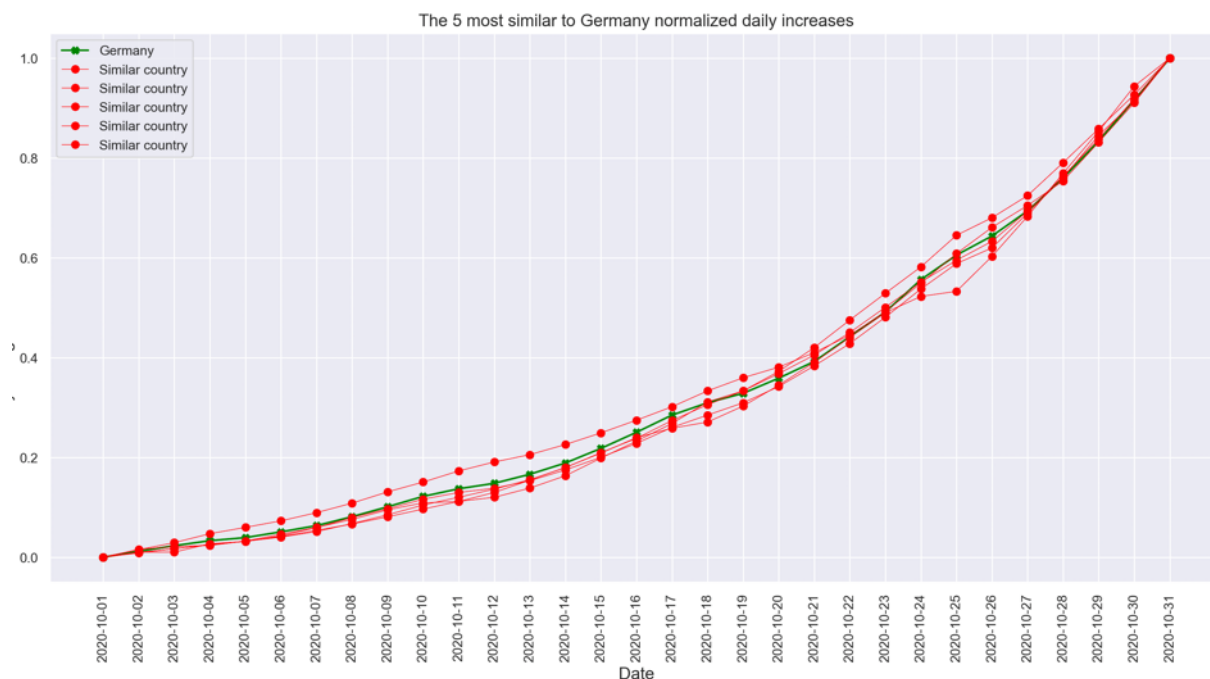


Rys. 53. Wykres pudełkowy przedstawiający w sposób graficzny reprezentację podstawowej analizy statystycznej kraju bazowego (tutaj Niemcy) oraz n-najbliższych sąsiadów. Pierwszy kraj to kraj bazowy, kolejne to jego najbliżsi sąsiedzi. Źródło: opracowanie własne.

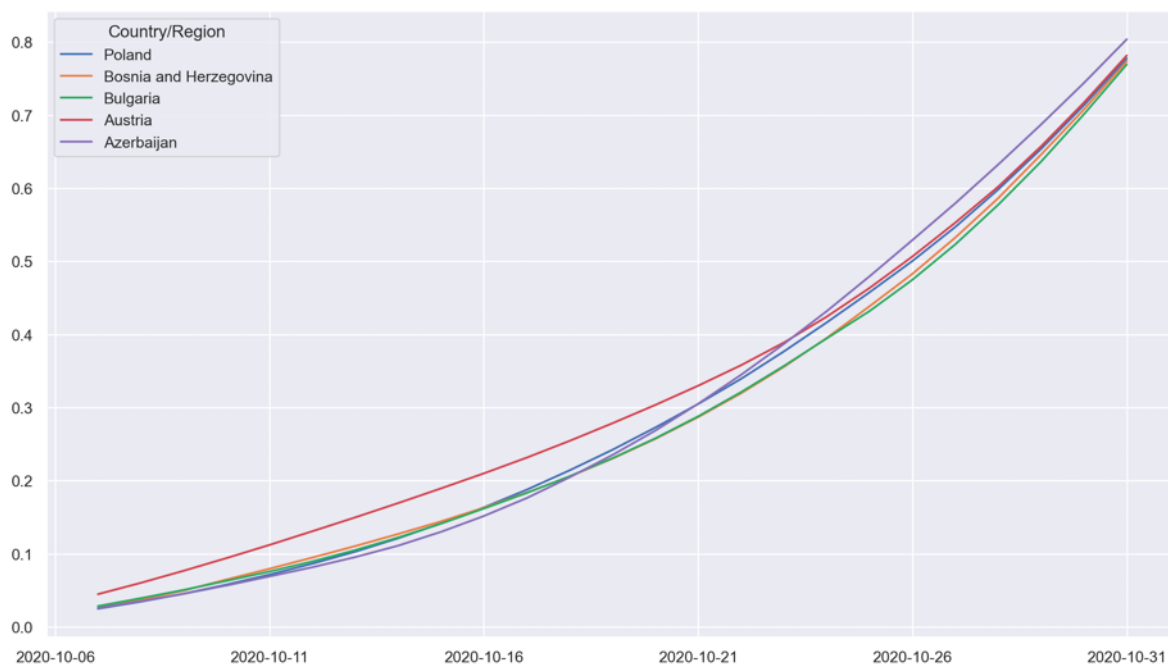
Ostateczna weryfikacja wyboru krajów odbywa się w postaci analizy wizualnej. Odbywa się to za pomocą zobrazowania przebiegów rzeczywistych i średniej kroczącej (7-dniowej) szeregów czasowych n-najbliższych sąsiadów. Rzeczywiste przebiegi n-najbliższych sąsiadów zobrazowano na Rys. 54. Zastosowano tutaj podejście ułatwiające obserwację w celu weryfikacyjnym poprzez „anonimizację”. Przebiegi każdego sąsiada przedstawione są w postaci linii łamanych oznaczone jednolitym kolorem (czerwonym) bez przypisania dla konkretnego z sąsiadów. Kolorem zielonym oznaczono przebieg kraju referencyjnego. Podejście takie umożliwia skoncentrowanie się obserwatora na celu analizy – bliskości przebiegu sąsiada względem zaznaczonego innym kolorem referencyjnego kraju. Zastosowanie unikalnego koloru dla każdego z sąsiadów utrudniało obserwację i czytelność, szczególnie dla liczby  $n > 9$  sąsiadów.

Przebiegi średniej kroczącej są zobrazowane na Rys. 55. Przebiegi te są „wygładzone” względem rzeczywistych przebiegów. Tutaj zdecydowano się na zastosowanie unikalnych kolorów i identyfikację krajów. Wykres jest pozbawiony części informacji – nie zawiera

pierwszego tygodnia na osi rzędnych. Wynika to ze specyfikacji liczenia średniej kroczącej. Na wykresie widać, że wszystkie wykresy mają zbliżone znormalizowane przebiegi. W zakresie dat 1-21 października 2020 r. wyróżnia się przebieg dla Austrii (5 najbliższy sąsiad), ma relatywnie większe przesunięcie na osi odciętych ku górze względem przebiegów pozostałych krajów, jednak w dalszej części musi być bliski przebiegowi dla kraju referencyjnego, ponieważ jego analiza graficzna na wykresie pudełkowym (Rys. 53. ) nie odbiega znacząco od wykresu dla referencyjnego kraju.

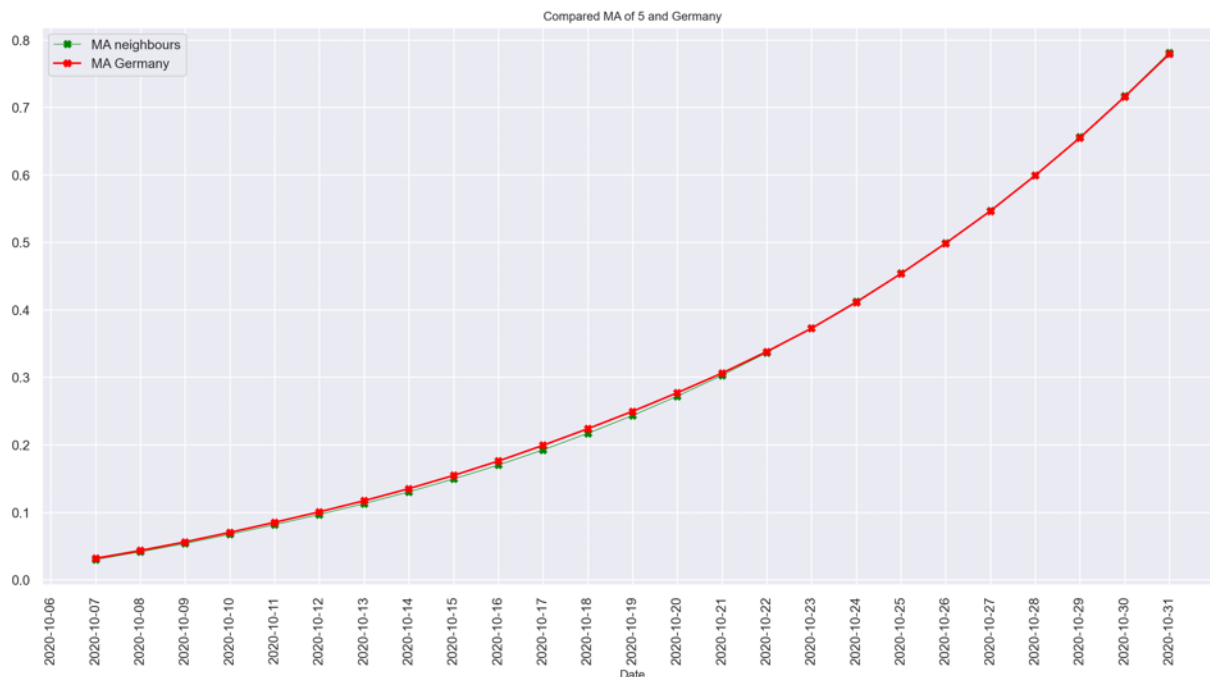


Rys. 54. Rzeczywiste znormalizowane przebiegi czasowe skumulowanej dobowej liczby zakażeń dla  $n$ -najbliższych sąsiadów (oznaczone kolorem czerwonym) oraz kraju referencyjnego (oznaczone kolorem zielonym). Źródło: opracowanie własne.



Rys. 55. Średnia krocząca MA znormalizowanych przebiegów szeregów czasowych skumulowanej dobowej liczby zakażeń dla  $n$ -najbliższych sąsiadów kraju referencyjnego. Źródło: opracowanie własne.

Otrzymane wartości średnich kroczących dla każdego kraju z  $n$ -sąsiadów są uśredniane, czego efektem jest jedna dobową wartość średnia. Zgodnie z założeniem metody, otrzymany przebieg powinien być zbliżony do referencyjnego, a w idealnym przypadku nakładać się na niego. Na Rys. 56. przedstawiono przebieg uśrednionej wartości szeregów czasowych  $n$ -sąsiadów i kraju referencyjnego, które niemal całkowicie nakładają się na siebie.



Rys. 56. Wykres przebiegów uśrednionej średniej kroczącej dla  $n$ -sąsiadów dobowej skumulowanej liczby zakażeń (oznaczona kolorem czerwonym) oraz średniej kroczącej skumulowanej dobowej liczby zakażeń dla kraju referencyjnego (oznaczona kolorem zielonym). Źródło: opracowanie własne.

Za pomocą równania (44) obliczane są wartości prognozowane szeregu czasowego. Dokładność predykcji wyniosła  $MAPE = 0,20$ . Zestawienie wyników w postaci numerycznej wraz z rzeczywistymi wartościami przedstawiono w Tab. 35. , natomiast graficznie na Rys. 57. Przebiegi MAPE i APE zawarto na Rys. 58. W pierwszym dniu prognozy zawsze zachodzi równość  $APE=MAPE$ .

Tab. 35. Zestawienie wartości rzeczywistych i prognozowanych oraz APE i MAPE dla badania predykcji o parametrach  $n=5$  i  $hp=3$  dla Niemiec. Źródło: opracowanie własne.

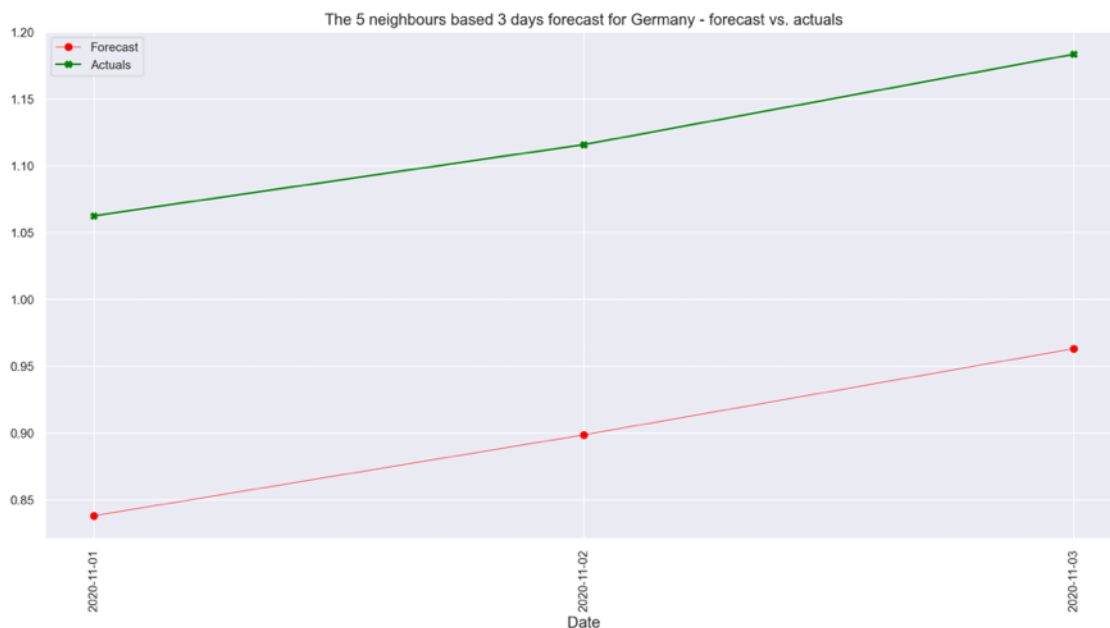
Data	Wartości prognozowane	Wartości rzeczywiste	APE	MAPE
1.11.2020	0,837	1,062	0,211	0,211
2.11.2020	0,898	1,116	0,194	0,204
3.11.2020	0,963	1,183	0,186	0,197

Na potrzeby graficznego zobrazowania przebiegów szeregów czasowych prognozowanych i rzeczywistych w zastosowanej metodzie stworzono wykres, który na krzywej do dnia początku horyzontu predykcji przedstawia dane rzeczywiste, a wartości w datach odpowiadających horyzontowi predykcji przedstawiają dane prognozowane (Rys. 57. ). Rozwiązanie to umożliwia wizualne określenie zachowania trendu wartości rzeczywistych.

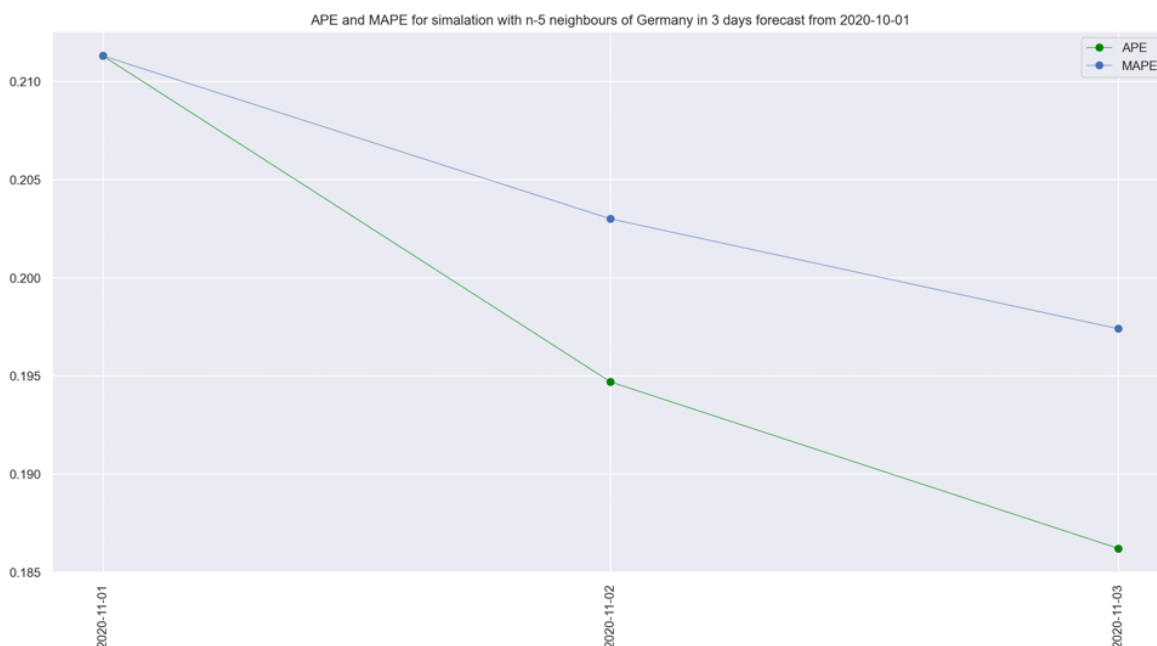


Rys. 57. Przebieg rzeczywisty (kolor zielony) oraz prognoza w horyzoncie predykcji (kolor niebieski) w symulacji o parametrach  $n=5$  i  $hp=3$  w badanym okresie czasu dla Niemiec. Źródło: opracowanie własne.

Następnie porównano na Rys. 58. wartości prognozowane i rzeczywiste oraz dokonano ewaluacji wyników na podstawie APE i MAPE (Rys. 59. ).

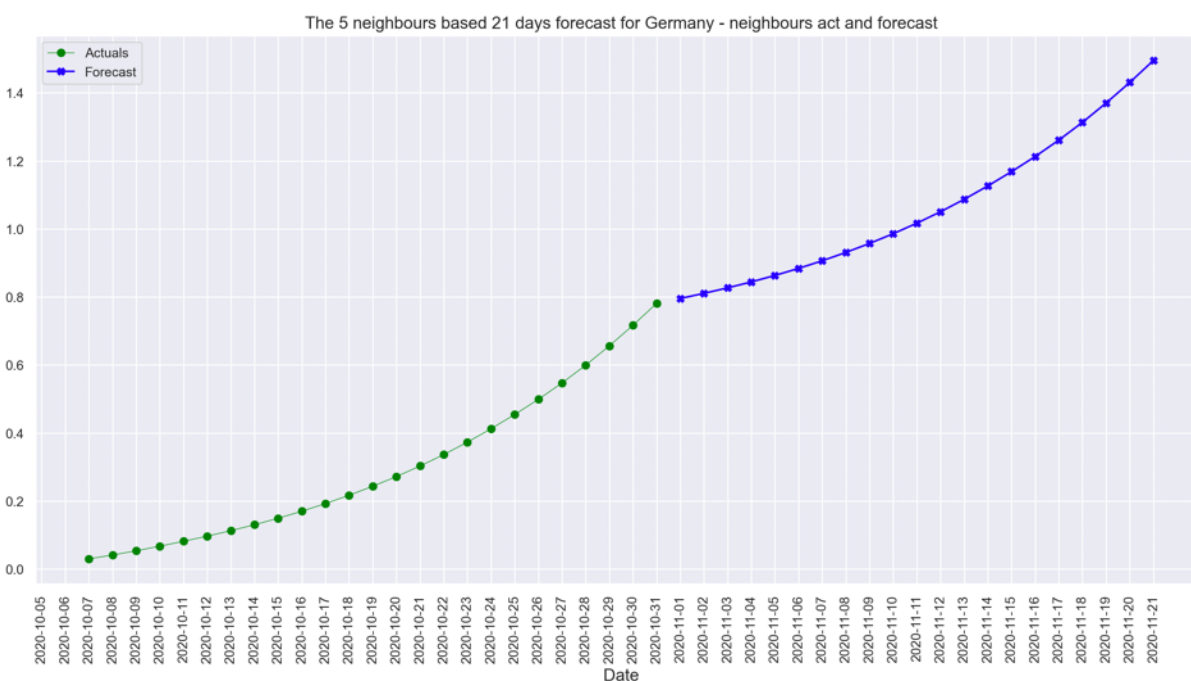


Rys. 58. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji  $h_p$ : rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Niemiec i parametrów  $n=5$  i  $h_p=3$ . Źródło: opracowanie własne.

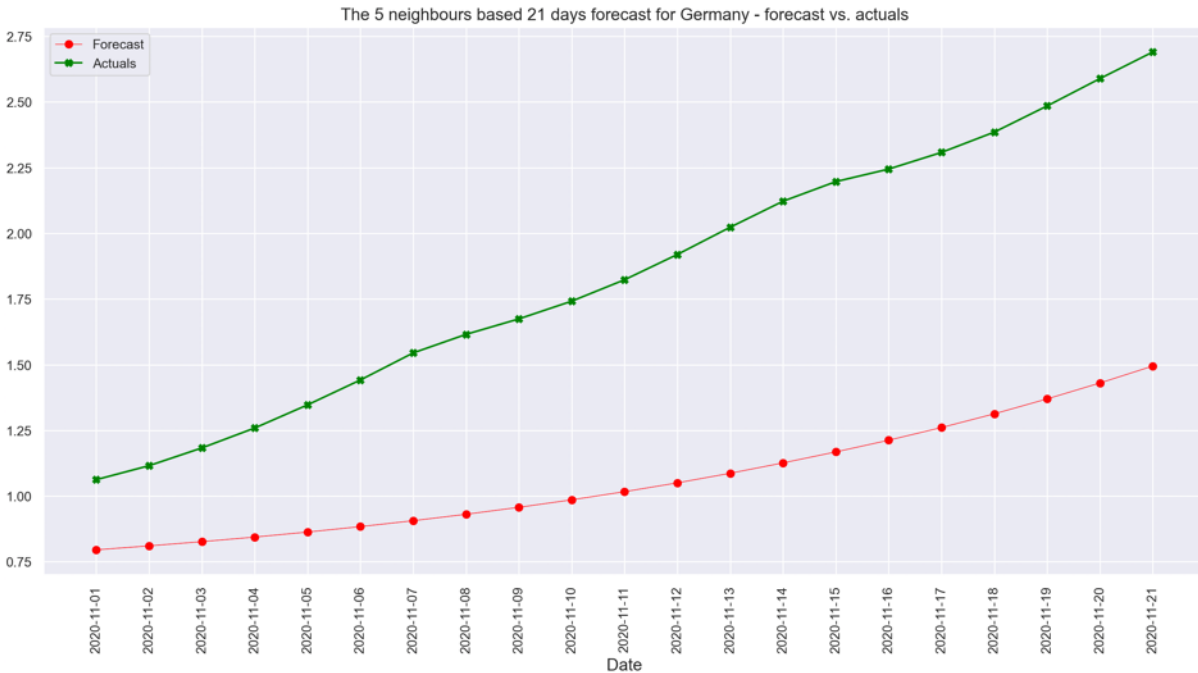


Rys. 59. Przebieg APE (kolor zielony) oraz MAPE (kolor niebieski) dla badanego przypadku – Niemiec, parametry  $n=5$  i  $h_p=3$ . Źródło: opracowanie własne.

Wykonano badania dla horyzontu predykcji  $hp=7$ ,  $hp=14$  i  $hp=21$ . Interesujące są wybrane wyniki i wykresy prognozy dla horyzontu predykcji  $hp=21$ . Metoda wizualnego przedstawienia wyników na wykresie zawierającym prognozowane dane zaraz po danych rzeczywistych pokazuje delikatne „złamanie” przebiegu krzywej pomiędzy obydwoma przebiegami (Rys. 60. ). Z kolei na Rys. 61. Przedstawiono rzeczywiste i prognozowane wartości. Rozwinięciem prac nad algorytmem może być znalezienie współczynników równania predykcji minimalizujących załamanie wykresu na styku przebiegu szeregów czasowych rzeczywistego i prognozowanego w taki sposób, aby prognozowany szereg stanowił przedłużenie trendu szeregu rzeczywistego.

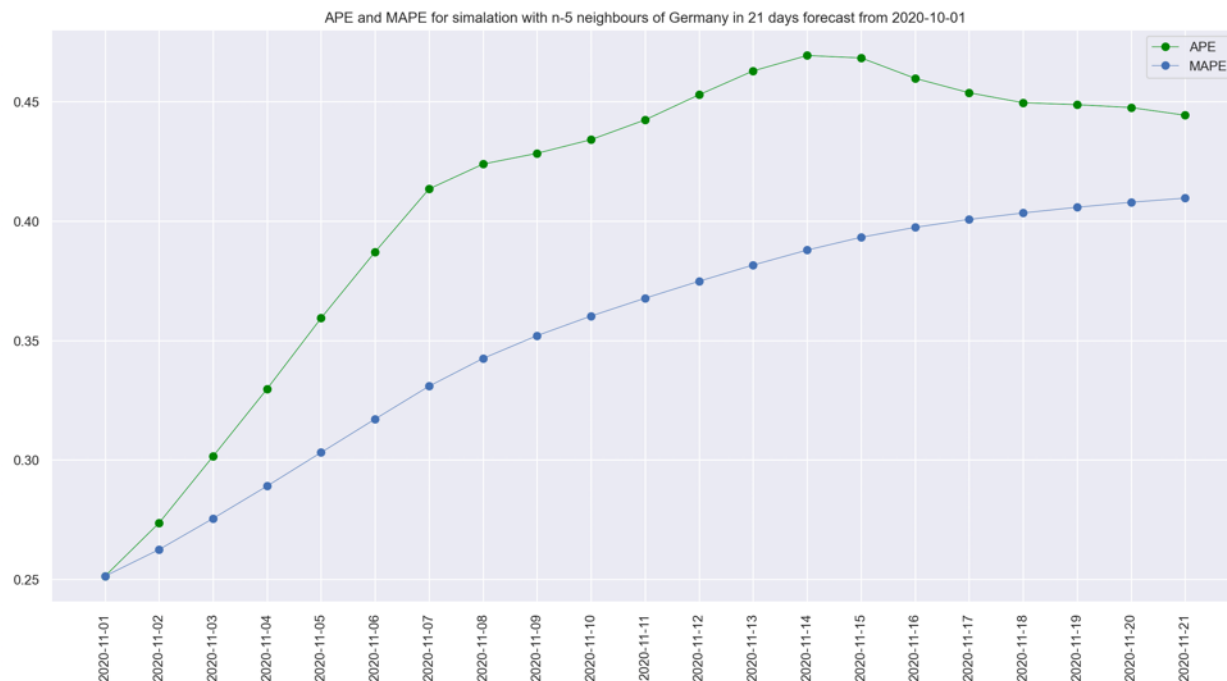


Rys. 60. Przebieg rzeczywisty (kolor zielony) oraz prognoza w horyzoncie predykcji (kolor niebieski) w symulacji o parametrach  $n=5$  i  $hp=21$  w badanym okresie czasu dla Niemiec. Źródło: opracowanie własne.



Rys. 61. Przebieg rzeczywisty (kolor zielony) oraz prognoza w horyzoncie predykcji (kolor niebieski) w symulacji o parametrach  $n=5$  i  $hp=21$  w badanym okresie czasu dla Niemiec. Źródło: opracowanie własne.

Graficzna analiza przebiegów MAPE i APE (Rys. 62. ) pozwala zauważyć, że wraz z kolejnym dniem w horyzoncie predykcji MAPE rośnie, przy czym po 13-tym dniu w horyzoncie predykcji APE zmniejsza się, co stawia metodę opartą na  $n$ -najbliższych sąsiadach jako obiecującą w przypadku relatywnie dłuższych horyzontów predykcji. MAPE dla przeprowadzonego badania zawiera się pomiędzy 0,25 a 0,41. Jest to wynik świadczący o wysokiej dokładności metody i zachęcający do dalszych badań i optymalizacji metody.



Rys. 62. Przebieg APE (kolor zielony) oraz MAPE (kolor niebieski) dla badanego przypadku – Niemiec, parametry  $n=5$  i  $hp=21$ . Źródło: opracowanie własne.

Kolejne badania powtórzono dla zmieniając liczbę  $n$ -najbliższych sąsiadów kolejno na 3, 7, 9 i 13. Zebrane wyniki predykcji w postaci macierzy zawierającej otrzymane wartości MAPE umieszczono w Tab. 36.

Tab. 36. Zbiorczy wynik MAPE przeprowadzonego badania metodą  $n$ -najbliższych sąsiadów z uwzględnieniem zmiany liczby sąsiadów i horyzontu predykcji. Źródło: opracowanie własne.

		n-najbliższych sąsiadów				
		3	5	7	9	13
Horyzont predykcji hp	3	0,20	0,20	0,19	0,19	0,19
	7	0,24	0,24	0,23	0,23	0,23
	14	0,33	0,32	0,31	0,32	0,32
	21	0,42	0,41	0,40	0,41	0,41

Tab. 37. Procentowe porównanie wyników MAPE względem wyników dla optymalnej wartości  $n=7$ . Źródło: opracowanie własne.

		Procentowa różnica n-najbliższych sąsiadów względem wyników dla $n=7$				
		3	5	7	9	13
Horyzont predykcji hp	3	7,5%	7,5%	0,0%	3,3%	3,6%
	7	7,3%	4,4%	0,0%	3,1%	2,9%
	14	5,2%	3,1%	0,0%	2,9%	2,9%
	21	3,8%	2,2%	0,0%	1,7%	2,4%

Niezależnie od długości horyzontu predykcji najmniejsze wartości MAPE uzyskano dla 7 n-najbliższych sąsiadów. Zwiększanie oraz zmniejszanie parametru n-najbliższych sąsiadów powoduje zwiększenie MAPE zgodnie z Tab. 37. W pozostałych badaniach przyjęto zatem parametr  $n=7$  jako optymalny do predykcji metodą n-najbliższych sąsiadów ze średnią krocząca.

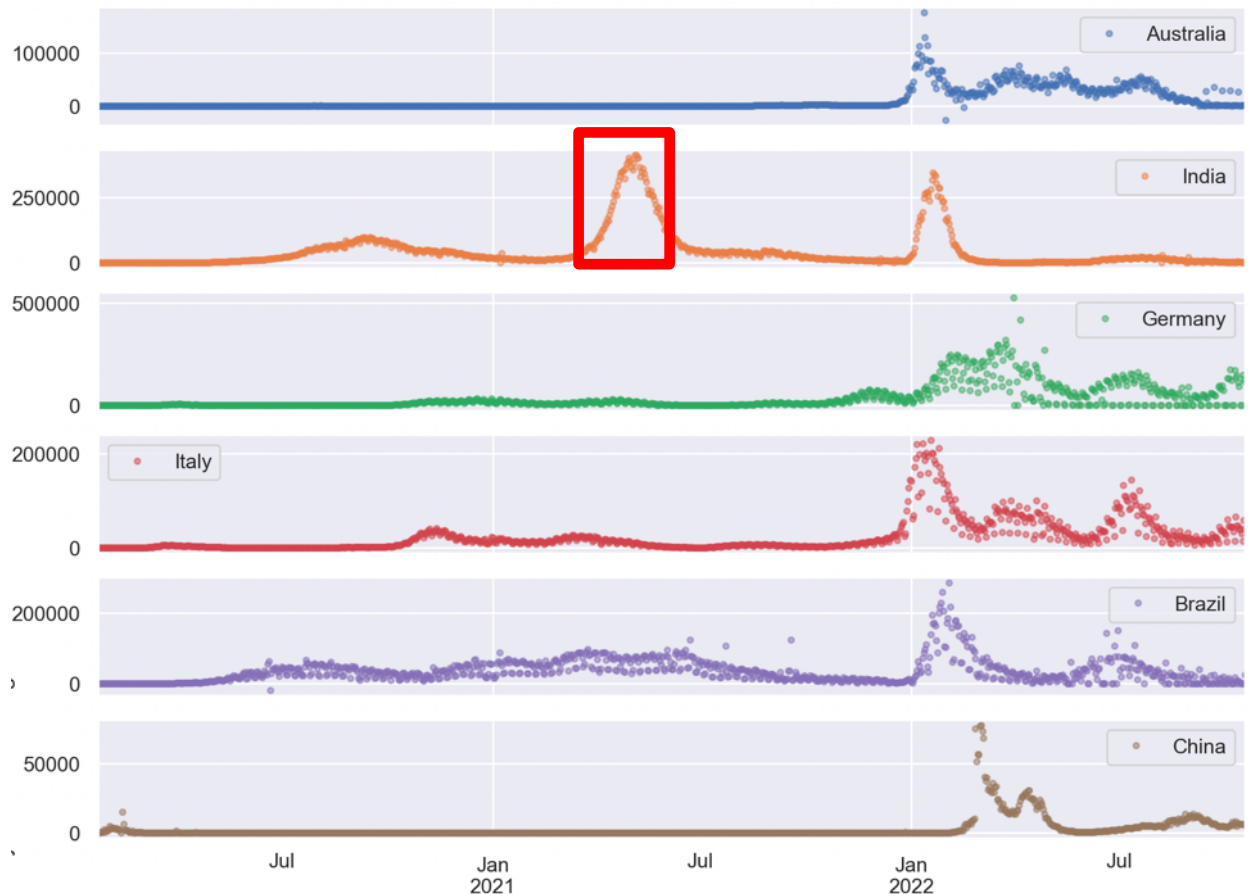
### 3.4 Badanie dokładności metod predykcji w tożsamym interwale czasowym w zależności od długości horyzontu predykcji.

Badanie umożliwia porównanie dokładności predykcji zakażeń za pomocą wszystkich opracowanych w tym celu metod i stanowi realizację zadań związanych z pytaniami Q1 i Q5 określonymi w planie prac badawczych w podrozdziale 3.1. Istotny jest wybór zakresu danych i kraju. Dane powinny obejmować okres początku fali zakażeń w najbardziej dotkniętym kraju. Prognozowanie na ekstremalnym przypadku danych będzie wyzwaniem dla metod, a dokładne predykcje potwierdzą ich przydatność w zastosowaniu w realnym środowisku. Badanie dokładności predykcji za pomocą każdej z metod połączono z badaniem dokładności predykcji wraz ze wzrostem horyzontu predykcji. Badanie dokładności predykcji w horyzoncie predykcji przekraczającym jeden tydzień jest w zasadzie możliwe z prezentowanymi metodami z wyłączeniem metody opartej na łańcuchach Markowa. Metoda oparta na łańcuchach Markowa

została opracowana do predykcji w horyzoncie predykcji  $hp=1$ ,  $hp=3$  i  $hp=7$ . Spowodowane jest to działaniem w tygodniowych oknach czasowych. Założono, że badana data 7 kwietnia 2021 r. musi zostać ujęta w oknie czasowym predykcji, kolejność dnia w oknie czasowym nie robi różnicy, liczy się przynależność do bina poprzedzającego elementu szeregu czasowego.

W przypadku pozostałych metod badanie zostanie wykonane w horyzoncie predykcji  $hp=7$ ,  $hp=14$  i  $hp=21$ . Predykcja dla  $hp=1$  i  $hp=3$  zostanie określona na podstawie pierwszych  $hp$  dni z predykcji  $hp=7$ .

Kolejnym zadaniem jest wybór kraju do badania. W tym celu wygenerowano wykres szeregów czasowych dziennych przyrostów zakażeń dla wybranych krajów (Rys. 62. ). Do porównania wybrano kraje, w których skutki zakażeń przyjmowały dramatyczne skutki nagłaśniane przez światowe media. Australia stosowała kontrowersyjne obostrzenia polegające na izolacji na granicy łamania praw człowieka. W Indiach wystąpił kryzys służby zdrowia w czasie lawinowego wzrostu zakażeń w okresie od maja do lipca 2021 r. Ludzie umierali przed szpitalami, zwłoki palono na osiedlach mieszkalnych. Niemcy wybrano do porównania jako jeden z największych krajów europejskich z rozwiniętą gospodarką. Analiza przebiegu szeregu czasowego dziennych wzrostów zakażeń (Rys. 63. ) pokazuje nieregularność w raportowaniu wyników, co może być spowodowane zaniżonymi danymi z dni wolnych od pracy, kiedy służba zdrowia pracuje w ograniczonym wymiarze. Podobny problem z danymi zaobserwowano w przypadku Włoch i Brazylii. Dane pochodzące z Chin nie zawierają wyraźnych zmian w przebiegu szeregu czasowego. Wybór padł na Indie, ze względu na wyraźne 3 fale.



Rys. 63. Wykres szeregów czasowych dziennych przyrostów zakażeń w wybranych krajach. Czerwona ramka oznacza badany okres dla wybranego kraju do badań. Źródło: opracowanie własne.

Wybrany czas rozpoczęcia okna predykcji to 7 kwietnia 2021 r. Jest to połowa fali wznoszącej liczby zakażeń. Wybór predykcji w tak ekstremalnym przypadku wzrostu zakażeń jest poważnym wyzwaniem badawczym.

### 3.4.1 Metoda multiregresji z pseudoinwersją Moore'a-Penrose'a.

Badanie metodą multiregresji z pseudoinwersją Moore'a-Penrose'a wykonano dla horyzontu predykcji  $hp=7$ ,  $hp=14$  i  $hp=21$ . Dokładność prognozy w horyzoncie predykcji  $hp=1$  i  $hp=3$  zbadano analizując predykcję w  $hp=7$ , wybierając MAPE odpowiednio dla 1 i 3 dnia w horyzoncie predykcji. Dla każdego badania i w celu osiągnięcia predykcji od 7 kwietnia 2021 r. w zadanym horyzoncie predykcji i uzyskaniu danych do ewaluacji wyników, zdefiniowano

odpowiednie parametry skryptu Python, takie jak szerokość okna i szerokość macierzy, a co za tym idzie zakres dat uczenia. Dla każdej zmiany parametry horyzontu predykcji algorytm musiał sięgać do wcześniejszych danych, ponieważ zwiększając horyzont predykcji należy rozszerzyć okno uczenia się, zatem data pomiaru gęstości szczepień i lista krajów również się zmienia.

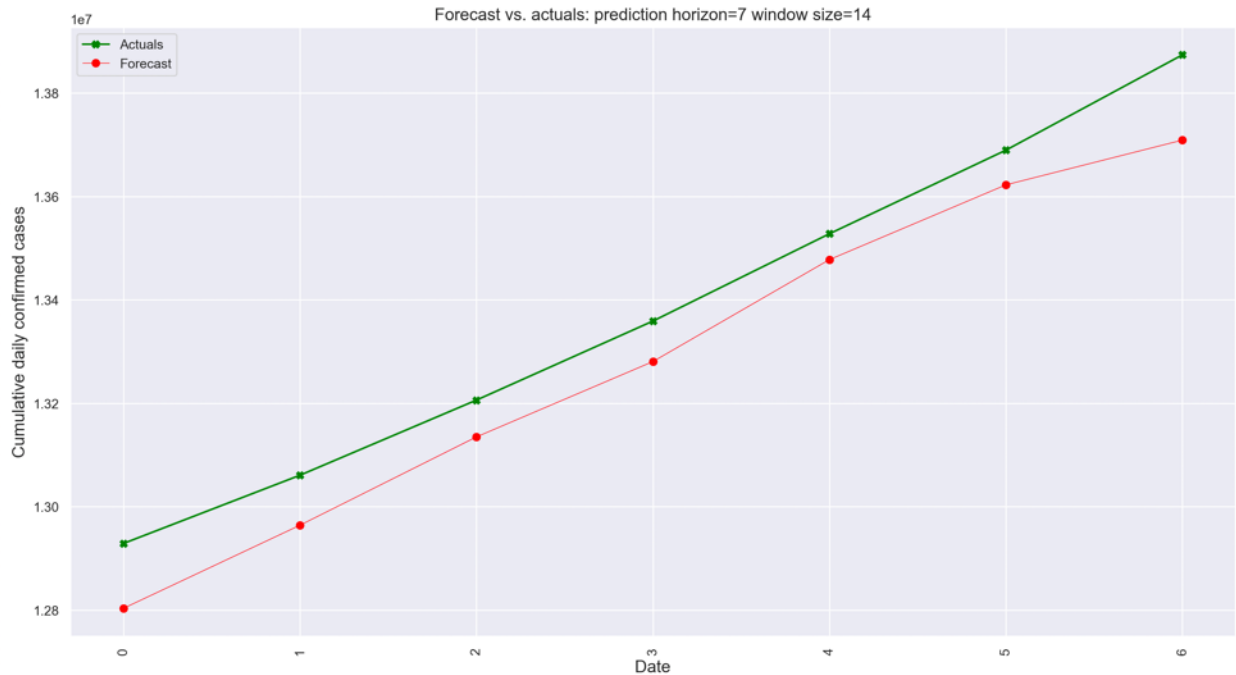
Przyjrzyjmy się bliżej wynikom predykcji  $hp=7$  (zawierającym dane w 3 badanych horyzontach predykcji) i  $hp=21$  z najdłuższym horyzontem predykcji.

Predykcja w zakresie  $hp=7$  okazała się bardzo dokładna (Rys. 64. ) osiągając wartości MAPE poniżej 0,01. Dla  $hp=1$  MAPE Najlepszy wynik osiągnięto dla  $hp=7$ , gdzie  $MAPE = 0,007$  (Tab. 37. )



Rys. 64. Przebieg APE dla predykcji szeregu czasowego metodą multiregresji z pseudo-inwersją Moore'a-Penrose'a dla  $hp=7$ . Wartość  $APE(0)$  odpowiada wartości MAPE dla  $hp=1$ . Źródło: opracowanie własne.

Graficzne zestawienie prognozowanych i rzeczywistych wartości szeregu czasowego (Rys. 65. ) potwierdza dokładność predykcji, oba wykresy mają niemal równoległe przebiegi.

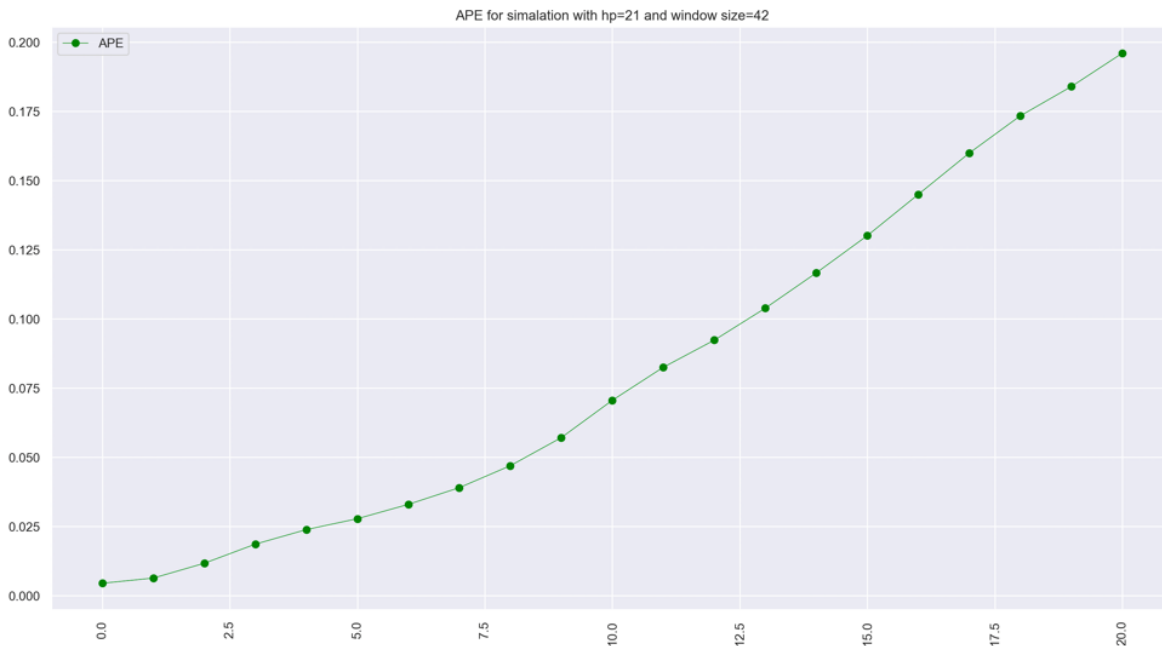


Rys. 65. Wartości rzeczywiste (kolor zielony) i prognozowane (kolor czerwony) szeregów czasowych - predykcja metodą multiregresji z pseudo-inwersją Moore'a-Penrose'a dla  $hp=7$ . Źródło: opracowanie własne.

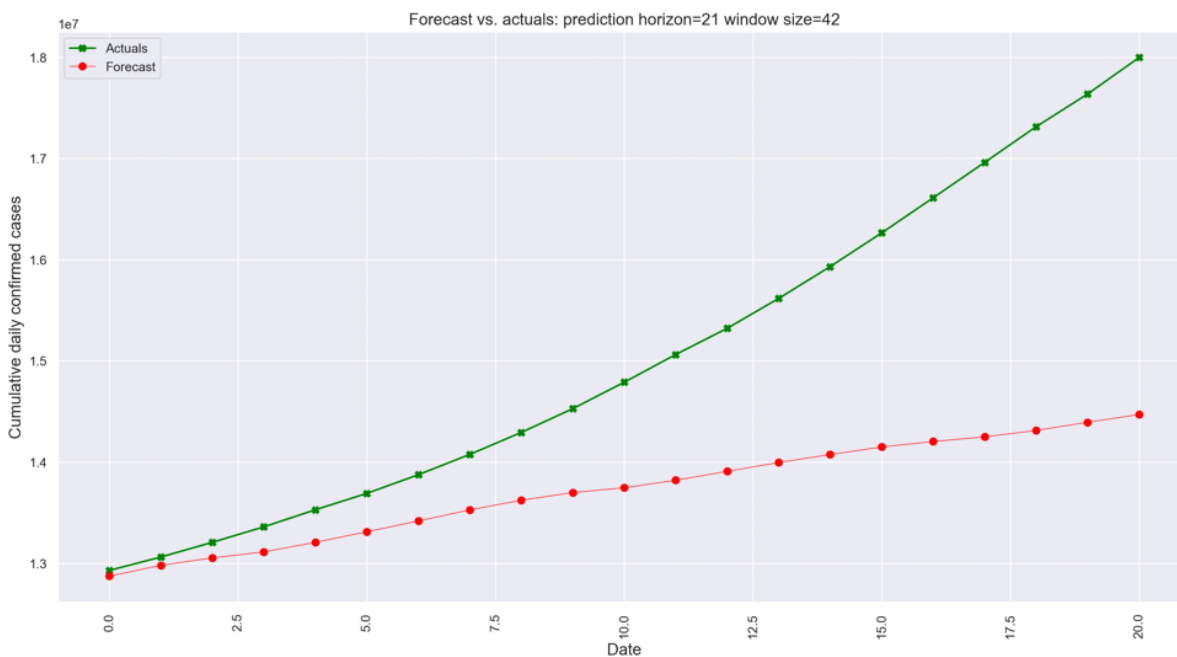
W przypadku prognozy z  $hp=21$  wystąpiło zwiększenie MAPE do wartości 0,082. Błąd APE rośnie wewnątrz horyzontu predykcji wraz z kolejnym dniem przyjmując przebieg zbliżony do liniowego (Tab. 38. , Rys. 66. ). Taki wzrost APE przekłada się odbieganiem od siebie wartości rzeczywistych i prognozowanych badanego szeregu czasowego (Rys. 67. ).

Tab. 38. Wynik predykcji szeregów czasowych skumulowanych dobowych zakażeń metodą multiregresji z pseudo-inwersją Moore'a-Penrose'a dla  $hp=21$ . Źródło: opracowanie własne.

Data	Wartości rzeczywiste	Wartości prognozowane	APE	MAPE
7.04.2021	12928574	12871459	0,004	0,004
8.04.2021	13060542	12978495	0,006	0,005
9.04.2021	13205926	13052124	0,012	0,007
10.04.2021	13358805	13111042	0,019	0,010
11.04.2021	13527717	13205678	0,024	0,013
12.04.2021	13689453	13309704	0,028	0,015
13.04.2021	13873825	13417102	0,033	0,018
14.04.2021	14074564	13526064	0,039	0,021
15.04.2021	14291917	13621651	0,047	0,023
16.04.2021	14526609	13697847	0,057	0,027
17.04.2021	14788003	13744841	0,071	0,031
18.04.2021	15061805	13819429	0,082	0,035
19.04.2021	15320972	13907019	0,092	0,040
20.04.2021	15616130	13994465	0,104	0,044
21.04.2021	15930774	14073143	0,117	0,049
22.04.2021	16263695	14148335	0,130	0,054
23.04.2021	16610481	14202342	0,145	0,059
24.04.2021	16960172	14247846	0,160	0,065
25.04.2021	17313163	14311645	0,173	0,071
26.04.2021	17636186	14392047	0,184	0,076
27.04.2021	17997113	14469219	0,196	0,082



Rys. 66. Przebieg APE dla predykcji szeregu czasowego metodą multiregresji z pseudo-inwersją Moore'a-Penrose'a dla  $hp=7$ . Źródło: opracowanie własne.



Rys. 67. Wartości rzeczywiste (kolor zielony) i prognozowane (kolor czerwony) szeregów czasowych - predykcja metodą multiregresji z pseudo-inwersją Moore'a-Penrose'a dla  $hp=21$ . Źródło: opracowanie własne.

Zestawienie wyników badania predykcji metodą multiregresji z pseudoinwersją Moore'a-Penrose'a na podstawie MAPE (Tab. 39. ) pokazuje, że najdokładniejszą predykcję osiągnięto w  $hp=7$ .

Tab. 39. Zbiornicze zestawienie wyników MAPE dla predykcji metodą multiregresji z pseudoinwersją Moore'a-Penrose'a. Źródło: opracowanie własne.

Dzienny horyzont predykcji	MAPE	Kraje o zbliżonej gęstości szczepień
$hp=1$	0,0097	Meksyk, Belize, Bangladesz, Rwanda, Indonezja, Jordan
$hp=3$	0,0075	
$hp=7$	0,0070	
$hp=14$	0,0388	Nepal, Oman, Grenada, Peru, Indonezja, Jordan
$hp=21$	0,0820	Nepal, Meksyk, Jordan, Indonezja, Panama, Ekwador.

Badanie dokładności predykcji w  $hp$  powyżej 7 dni, czyli  $hp=14$  i  $hp=21$  wykazały wzrost MAPE o wielokrotności odpowiednio  $\sim 4,5$  i  $\sim 10,75$  względem MAPE( $hp=7$ ).

### 3.4.2 Metoda oparta na $n$ -najbliższych sąsiadach i średniej kroczącej.

Badanie oparto na  $n=7$   $n$ -najbliższych sąsiadów, ponieważ przy takiej wartości parametru  $n$  uzyskano najdokładniejsze prognozy. W przypadku tej metody zastosowano długość okna uczenia wynoszącą 42 dni. Krajami podobnymi w znormalizowanych przebiegach czasowych skumulowanej dobowej liczby zakażeń okazały się: Trynidad i Tabago, Azerbejdżan, Uzbekistan, Turcja, Filipiny, Mali i Papua Nowa Gwinea. W przeciwieństwie do metody multiregresji z pseudoinwersją Moore'a-Penrose'a, długość okna jest stała dla badanego zakresu  $hp$  i nie trzeba dostosowywać zakresu dat okna przed symulacją.

Otrzymane wyniki predykcji przedstawiono w Tab. 40.

Tab. 40. Zbiorcze zestawienie wyników MAPE dla predykcji metodą  $n$ -najbliższych sąsiadów. Źródło: opracowanie własne.

Dzienny horyzont predykcji	MAPE	Kraje o zbliżonych znormalizowanych przebiegach
hp=1	0,16	Trynidad i Tabago,
hp=3	0,18	Azerbejdżan,
hp=7	0,22	Uzbekistan, Turcja,
hp=14	0,32	Filipiny, Mali,
hp=21	0,44	Papua Nowa Gwinea

Podobnie jak w przypadku metody multiregresji z pseudoinwersji Moore'a-Penrose'a, bliżej przedstawiono wyniki predykcji badania w najdłuższym badanym horyzoncie predykcji hp=21 (Tab. 41. ).

Tab. 41. Wynik predykcji szeregów czasowych skumulowanych dobowych zakażeń metodą  $n$ -najbliższych sąsiadów dla  $hp=21$ . Źródło: opracowanie własne.

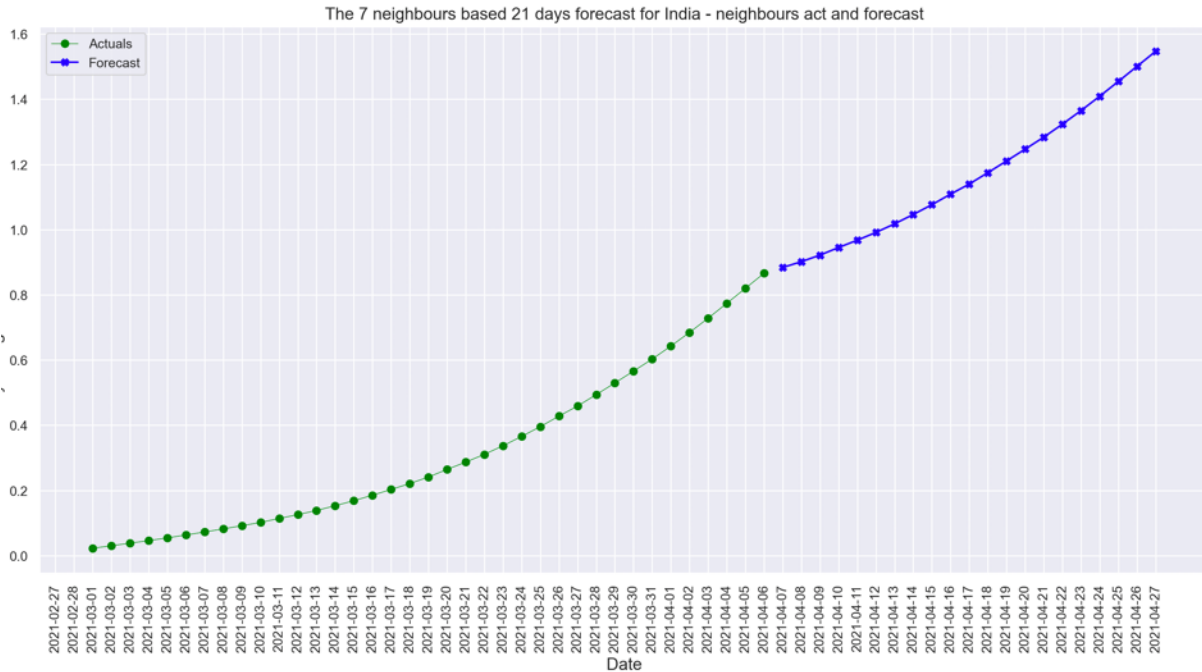
Data	Wartości rzeczywiste	Wartości prognozowane	APE	MAPE
7.04.2021	1,072	0,884	0,18	0,18
8.04.2021	1,146	0,902	0,21	0,19
9.04.2021	1,228	0,922	0,25	0,21
10.04.2021	1,314	0,945	0,28	0,23
11.04.2021	1,410	0,968	0,31	0,25
12.04.2021	1,501	0,991	0,34	0,26
13.04.2021	1,605	1,018	0,37	0,28
14.04.2021	1,718	1,046	0,39	0,29
15.04.2021	1,841	1,076	0,42	0,31
16.04.2021	1,974	1,109	0,44	0,32
17.04.2021	2,121	1,140	0,46	0,33
18.04.2021	2,276	1,175	0,48	0,34
19.04.2021	2,422	1,210	0,50	0,36
20.04.2021	2,589	1,247	0,52	0,37
21.04.2021	2,766	1,284	0,54	0,38
22.04.2021	2,954	1,323	0,55	0,39
23.04.2021	3,150	1,365	0,57	0,40
24.04.2021	3,347	1,409	0,58	0,41
25.04.2021	3,546	1,454	0,59	0,42
26.04.2021	3,729	1,501	0,60	0,43
27.04.2021	3,933	1,547	0,61	0,44

Przebiegi APE i MAPE zobrazowane na Rys. 68. pokazują stabilny wzrost badanych wielkości wraz ze wzrostem horyzontu predykcji. Przechodząc do wykresu z porównaniem wartości rzeczywistych i prognozowanych szeregów czasowych, widoczne jest zaniżenie wartości

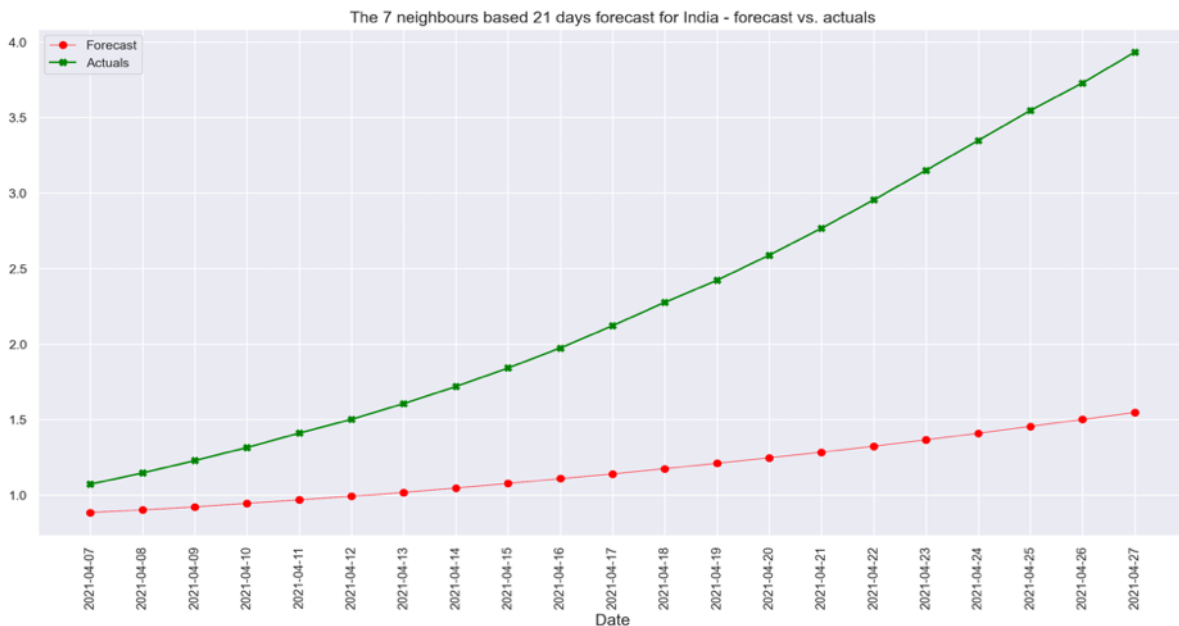
prognozowanej względem rzeczywistej (Rys. 69. ). W omawianym ekstremalnym przypadku fali wznoszącej pandemii w Indiach, zniżenie wartości prognozowanych byłby sygnałem ostrzegawczym dla władz kraju. Wynik predykcji zobrazowany również na Rys. 70. nie oddał w pełni skali skumulowanych zakażeń i problemu czekającego tamtejszą służbę zdrowia, jaki przedstawiają dane rzeczywiste.



Rys. 68. Przebieg APE (kolor zielony) i MAPE (kolor czerwony) dla predykcji szeregu czasowego metodą  $n$ -najbliższych sąsiadów dla  $hp=21$ . Źródło: opracowanie własne.



Rys. 69. Przebieg rzeczywisty (kolor zielony) oraz prognoza w horyzoncie predykcji (kolor niebieski) szeregów czasowych skumulowanych dobowych zakażeń - predykcja metodą n-najbliższych sąsiadów dla  $hp=21$ . Źródło: opracowanie własne.



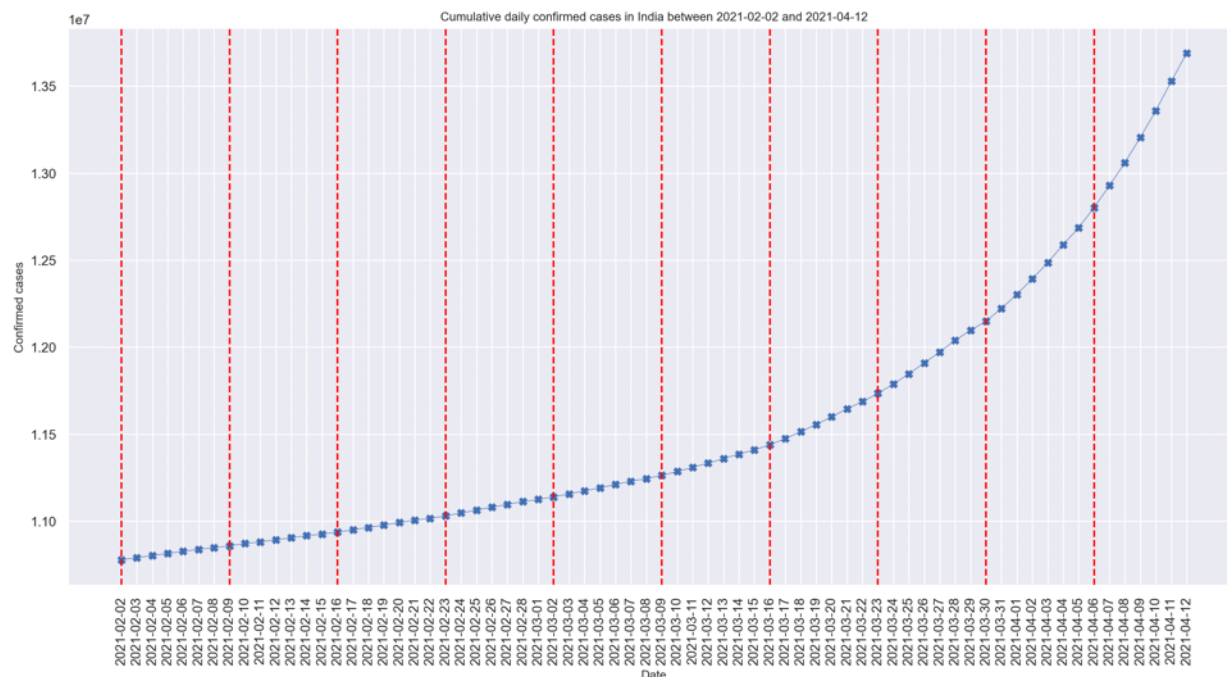
Rys. 70. Wartości rzeczywiste (kolor zielony) i prognozowane (kolor czerwony) szeregów czasowych skumulowanych dobowych zakażeń - predykcja metodą n-najbliższych sąsiadów dla  $hp=21$ . Źródło: opracowanie własne.

### 3.4.3 Metoda oparta na łańcuchach Markowa.

Badanie metodą opartą na łańcuchach Markowa wykonano z wykorzystaniem najdłuższego okna czasowego zastosowanego w poprzednich metodach.

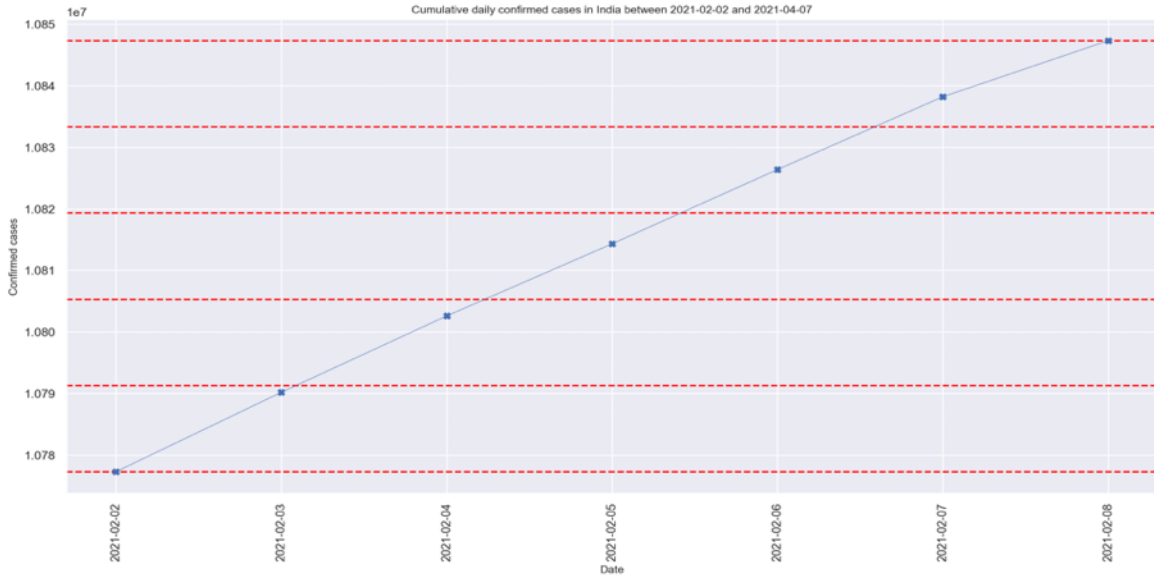
Istotną różnicą względem poprzednich metod jest predykcja wartości przyrostów dobowych zakażeń w tej metodzie, w poprzednich badana jest skumulowana wartość dobowych zakażeń.

Metoda ta ze względu na swoją specyfikę bada predykcję w tygodniowych oknach czasowych (Rys. 71. ). Istotne było umieszczenie daty początku horyzontu predykcji w ostatnim oknie czasowym., aby wykorzystać informację zgromadzoną w czasie uczenia w macierzach przejść  $M$  i powstałej na jej podstawie macierzy prawdopodobieństwa  $P$ .



Rys. 71. Przebieg skumulowanej liczby zakażeń w Indiach w zakresie dat uczenia z podziałem na tygodniowe okna. Źródło: opracowanie własne.

Dla każdego okna czasowego dokonano podziału na biny (przedziały wartości pomiędzy maksymalną i minimalną wartością w oknie) i sklasyfikowano każdą wartość szeregu czasowego w odpowiednim binie (Rys. 72. ).



Rys. 72. Podział pierwszego tygodniowego okna czasowego badanego szeregu czasowego skumulowanych dobowych zakażeń w Indiach na 5 binów. Źródło: opracowanie własne.

Przejścia pomiędzy binem źródłowym i docelowym wraz z czasem zostały zliczone i przedstawione w macierzy M (Tab. 42. ).

Tab. 42. Macierz przejść dla badanego szeregu czasowego. Źródło: opracowanie własne.

$$M = \begin{bmatrix} 10 & 10 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 2 & 10 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}$$

Na podstawie macierzy M została utworzona macierz prawdopodobieństwa P (Tab. 43. ), której maksymalna wartość w kolumnie określa numer bina docelowego. W Tab. 44. przedstawiono rzeczywiste wartości binów źródłowych, docelowych oraz prognozowane biny docelowe.

Tab. 43. Zobrazowana macierz prawdopodobieństw  $P$ . Źródło: opracowanie własne.

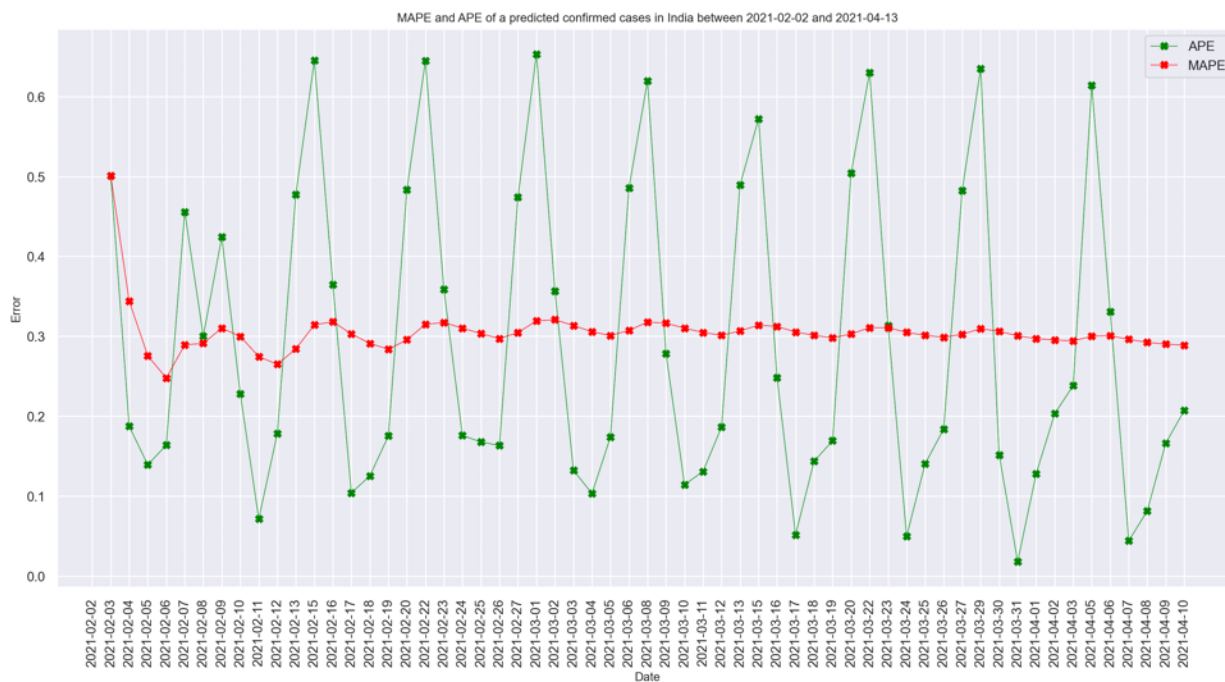
$$P = \begin{bmatrix} 0,5 & 0,5 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0,17 & 0,83 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Tab. 44. Wyniki analizy badanego okresu metodą z wykorzystaniem łańcuchów Markowa. Kolumny oznaczają kolejno: data, wartości rzeczywiste, wartości prognozowane, wartość odpowiadająca medianie bina, numer okna czasowego, numer bina źródłowego, docelowy bin, prognozowany docelowy bin, APE, MAPE. Źródło: opracowanie własne.

Data	Wartości rzeczywiste	Wartości prognozowane	Mediana bina	Nr okna czasowego	Bin źródłowy	Bin docelowy	Prognozowany bin docelowy	APE	MAPE
3.02.2021	12899	6441,84	7002	1	1	2	1	0,50	
4.02.2021	12408	10082,88	7002	1	2	3	3	0,19	
5.02.2021	11713	10082,88	7002	1	3	4	4	0,14	
6.02.2021	12059	10082,88	7002	1	4	5	5	0,16	
7.02.2021	11831	6441,84	7002	1	5	5	5	0,46	
8.02.2021	9110	6373,76	6928	2	1	1	1	0,30	
...									
3.04.2021	93249	71022,24	49321	9	4	5	5	0,24	
5.04.2021	200540	77433,64	84167	10	1	1	1	0,61	0,61
6.04.2021	115736	77433,64	84167	10	1	2	1	0,33	0,47
7.04.2021	126789	121200,48	84167	10	2	3	3	0,04	0,33
8.04.2021	131968	121200,48	84167	10	3	4	4	0,08	0,27
9.04.2021	145384	121200,48	84167	10	4	4	5	0,17	0,25
10.04.2021	152879	121200,48	84167	10	4	5	5	0,21	0,24

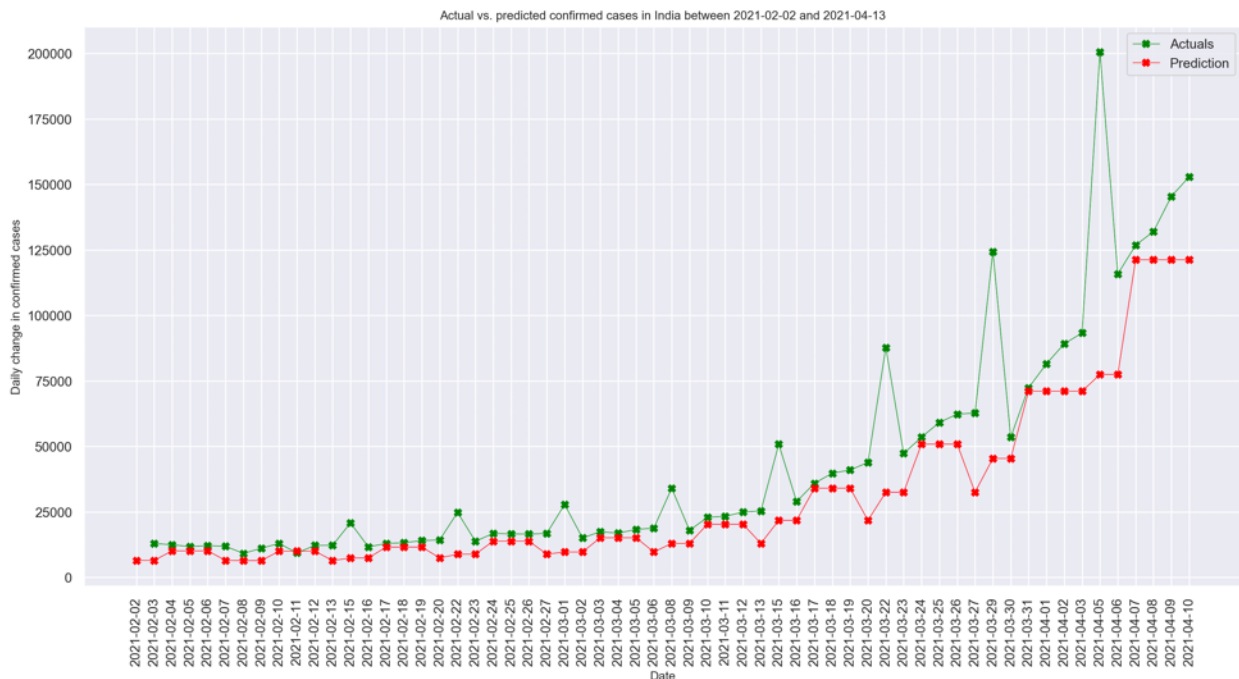
MAPE na potrzeby badania określono tylko i wyłącznie dla okna czasowego nr 10, w którym zawiera się badana data. MAPE jest liczone od daty w  $hp=1$ , to jest pierwszy dzień okna czasowego przypadający na 5 kwietnia, w innych metodach był to 7 kwietnia. Nie ma to znaczenia w przypadku tej metody, ponieważ dla całego okna obowiązuje ta sama macierz  $P$  określająca przejścia, istotne jest zawarcie badanego dnia początku  $hp$  w oknie czasowym. Wartość prognozowaną dla  $hp=7$  przyjęto ostatni dzień okna czasowego. Wykres na Rys. 73. Przedstawi MAPE i APE dla predykcji wraz z oknem uczenia.

Graficzna analiza MAPE pokazuje wartość MAPE w  $hp=1$  począwszy od okresu uczenia, nie jest to zatem MAPE uwzględnione w wynikach w Tab. 44. W pierwszym oknie czasowym, na etapie uczenia zaobserwowano wyższe wartości MAPE. Dla następnych wartości prognozowanych MAPE oscyluje z malejącą amplitudą, co jest wynikiem uwzględnienia uczenia z poprzednich okien czasowych, przyjmując wartość  $\sim 0,3$ . Wartości APE natomiast oscylują w zakresie  $\sim 0,018$  do  $\sim 0,65$ . Widać pewną okresowość zmian, który może być zinterpretowany jako wpływ raportowania danych epidemiologicznych w dni wolne od pracy, który był zaobserwowany w wielu publikacjach, w tym przy badaniu casu od infekcji do śmierci (Wiliński et al. (2022)).



Rys. 73. Przebieg APE (kolor zielony) i MAPE (kolor czerwony) w badanym okresie predykcji dla Indii. Źródło: opracowanie własne.

Badanie oparte na krzywej szeregu czasowego skumulowanych zdobytych zakażeń w omawianym przypadku fali wznoszącej Indii pozwala zaobserwować wady i zalety tej metody. Krzywą wartości rzeczywistych można podzielić na 2 fazy w zależności od dynamiki wzrostu wartości dobowych zakażeń. Za datę graniczną można przyjąć 17 marca, bowiem od tego dnia wartości rzeczywiste dobowych przyrostów zdecydowanie wzrastają względem wartości sprzed wspomnianej daty.



Rys. 74. Porównanie wartości rzeczywistych i predykcji dobowej liczby zakażeń dla Indii w badanym zakresie dat z uwzględnieniem ograniczeń metody. Kolorem zielonym oznaczono dane rzeczywiste, kolorem czerwonym wartości prognozowane. Źródło: opracowanie własne.

W przypadku mniejszej dynamiki wzrostów w pierwszej fazie wartości rzeczywiste i prognozowane różnią się nieznacznie na Rys. 74. Natomiast w drugiej fazie zdecydowanych wzrostów wartości dobowych szeregu czasowego można zaobserwować znaczne rozbieżności pomiędzy wartościami rzeczywistymi i prognozowanymi, zwłaszcza w tygodniu zaczynającym się od 31 marca a kończącym 6 kwietnia. Oznacza to, że zaletą tej metody jest relatywnie dobra dokładność predykcji jednodobowych przy umiarkowanych wzrostach dobowych zakażeń,

natomiast wadą jest zdecydowanie niższa dokładność predykcji przy ekstremalnym wzroście wartości dobowych zakażeń.

#### 3.4.4 Zbiorcze porównanie metod predykcji w zależności od horyzontu redykcji pod względem MAPE.

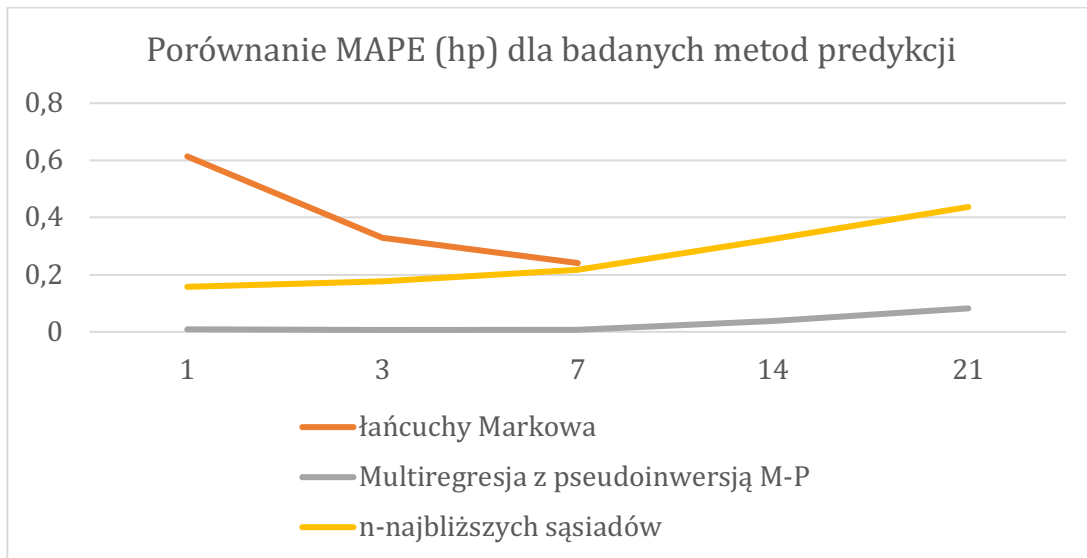
Przeprowadzone badania wyłoniły najdokładniejszą metodę predykcji, którą okazała się metoda multiregresji z pseudoinwersją Moore'a – Penrose'a. W przypadku  $hp=1$ ,  $hp=3$  i  $hp=7$  wartość MAPE zdecydowanie jest niższa – lepsza od pozostałych metod (Tab. 45. ). Najmniej dokładną metodą w  $hp=1$  okazała się metoda oparta na łańcuchach Markowa. Metoda ta osiągnęła zbliżony wynik do metody n-najbliższych sąsiadów w  $hp=7$ .

W przypadku metod opartych na n-najbliższych sąsiadach oraz multiregresji z pseudoinwersją Moore'a-Penrose'a MAPE rośnie wraz ze wzrostem horyzontu predykcji (Rys. 75. ), przy czym wzrosty są większe powyżej  $hp=7$ . W przypadku metody opartej na łańcuchach Markowa i obserwacji z Rys. 75. dokładność predykcji rośnie wraz ze wzrostem horyzontu predykcji. Jednak w przypadku tej metody należy pamiętać, że działa ona na podstawie tygodniowych okien czasowych, prognozuje dobowe przyrosty zamiast skumulowanych wartości oraz bazuje na jednej zmiennej.

Prognozy w najdłuższym badanym horyzoncie predykcji  $hp=21$  okazały się bardzo dokładne w przypadku multiregresji uzyskując imponującą wartość  $\sim 0,08$ , natomiast metoda n-najbliższych sąsiadów okazała się  $\sim 5,8$  razy mniej dokładna.

Tab. 45. Porównanie dokładności metod predykcji za pomocą MAPE w zależności od długości horyzontu predykcji. Źródło: opracowanie własne.

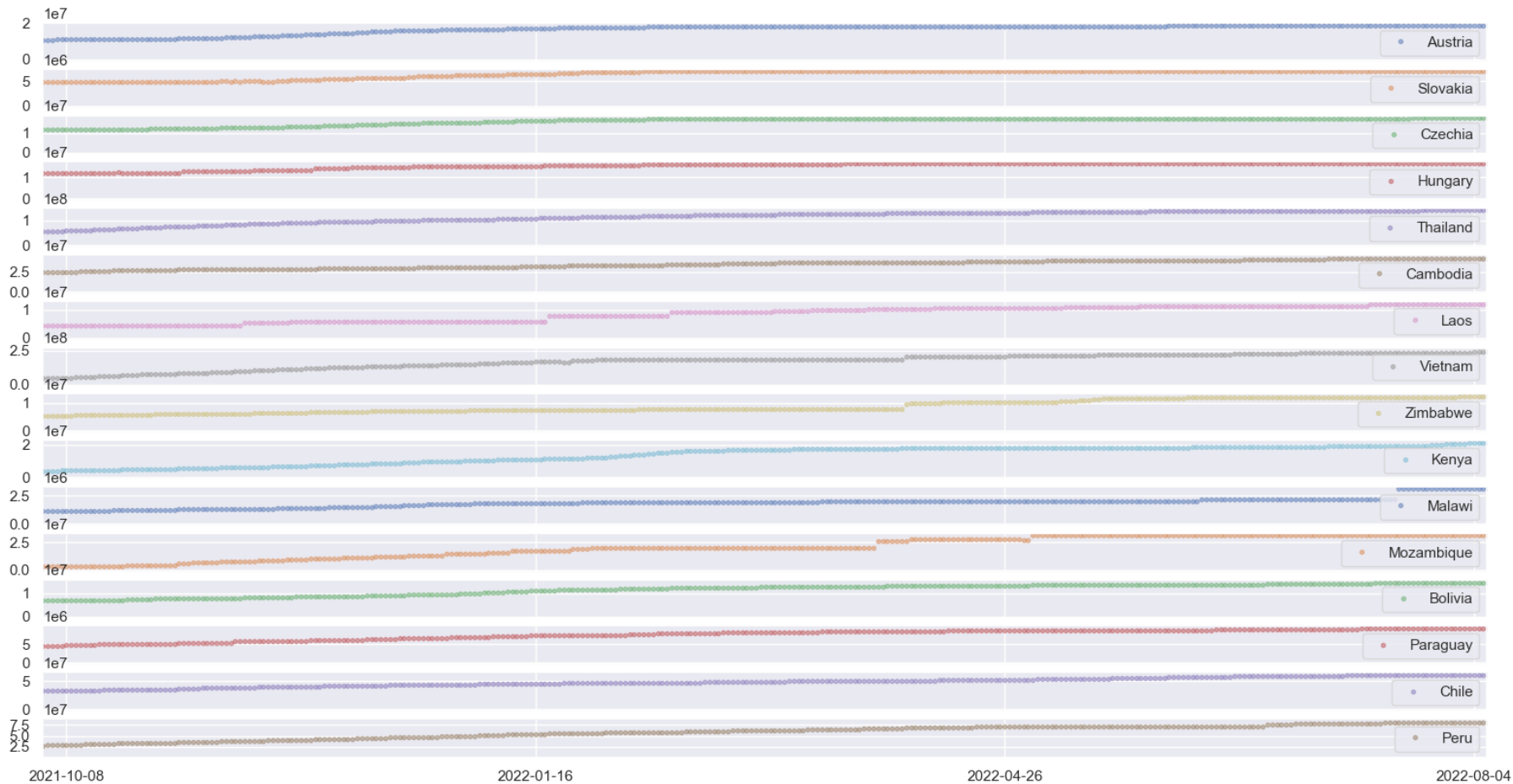
Dzienny horyzont predykcji	łańcuchy Markowa	Multiregresja z pseudoinwersją Moore'a-Penrose'a		n-najbliższych sąsiadów	
	MAPE	MAPE	Kraje o zbliżonej gęstości szczepień	MAPE	Kraje o zbliżonych znormalizowanych przebiegach
hp=1	0,61	0,0097	Meksyk, Belize,	0,16	Trynidad i Tabago, Azerbejdżan, Uzbekistan, Turcja, Filipiny, Mali, Papua Nowa Gwinea
hp=3	0,33	0,0075	Bangladesz, Rwanda,	0,18	
hp=7	0,24	0,0070	Indonezja, Jordan	0,22	
hp=14	-	0,039	Nepal, Oman, Grenada, Peru, Indonezja, Jordan	0,32	
hp=21	-	0,082	Nepal, Meksyk, Jordan, Indonezja, Panama, Ekwador.	0,44	



Rys. 75. Porównanie badanych metod pod względem dokładności predykcji szeregów czasowych w zależności od długości horyzontu predykcji hp za pomocą kryterium MAPE. Źródło: opracowanie własne.

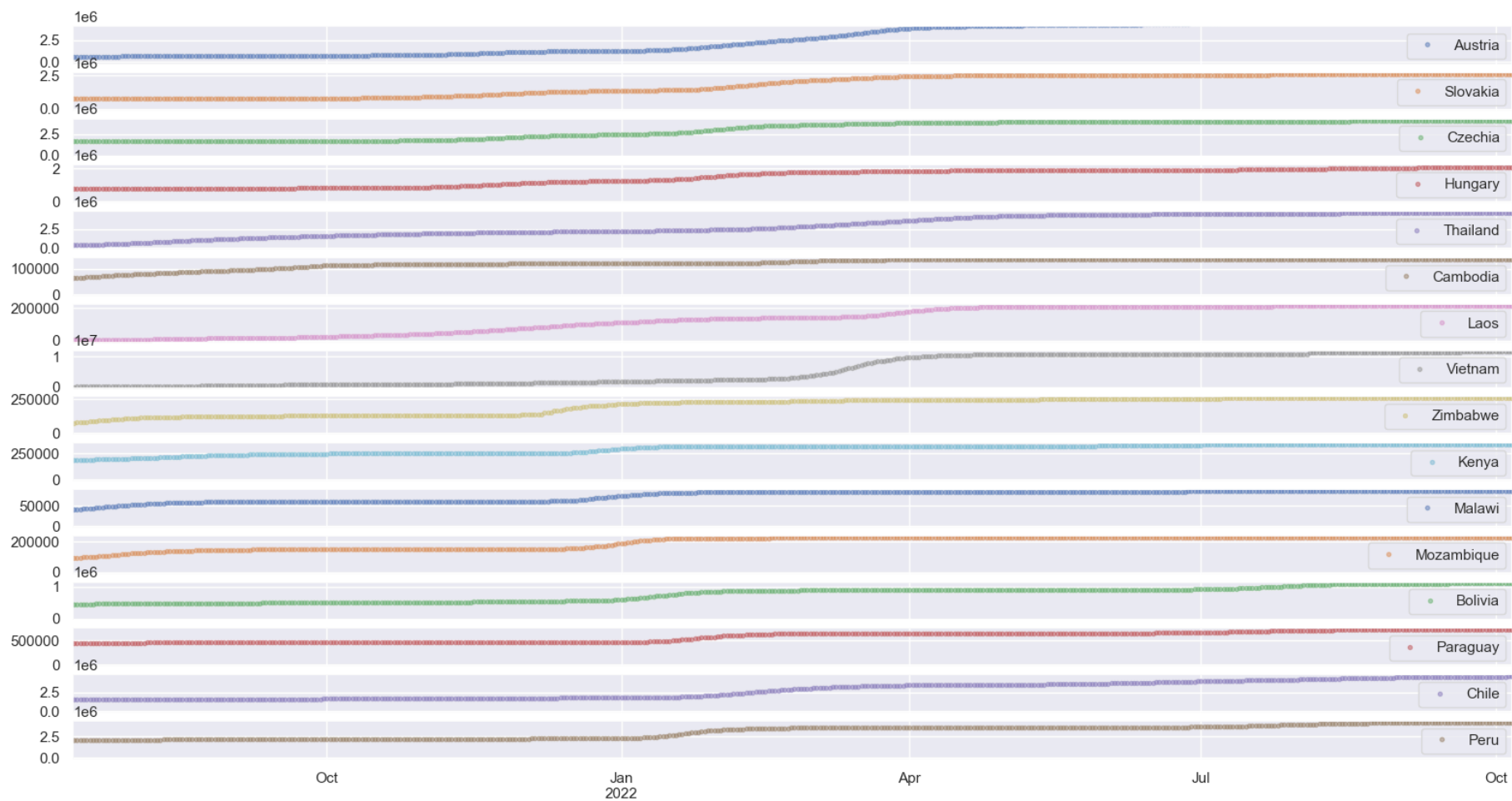
### 3.5 Wpływ położenia geograficznego na dokładność predykcji.

Jednym z czynników mającego wpływ w procesie złożonym jakim jest pandemia jest położenie geograficzne oraz związany z nim szereg innych czynników: kulturowych, klimatycznych, rozwój gospodarczy, etc. Badanie wpływu położenia geograficznego na dokładność predykcji stanowi realizację zadań związanych z pytaniem Q3 określonym w planie prac badawczych w podrozdziale 3.1. Badanie przeprowadzono w horyzoncie predykcji  $hp=7$ . Z 4 kontynentów wybrano po 4 kraje, zróżnicowane gospodarczo, klimatycznie i kulturowo. Warunkiem było również bliskie położenie krajów względem siebie oraz zbliżona powierzchnia. Nie udało się wypełnić początkowego założenia wyboru krajów sąsiadujących ze względu na brakujące dane, głównie dotyczące szczepień. W Europie uwzględniono Austrię, Słowację, Czechy i Węgry. W przypadku Azji wybór padł na Tajlandię, Laos, Wietnam i Kambodżę. Wśród krajów afrykańskich wybrano Zimbabwe, Mali, Mozambik i Kenię. Amerykę południową reprezentowały Peru, Chile, Boliwia i Paragwaj. Następnie określono zakres dat stanowiących horyzont predykcji. Daty te oraz poprzedzający okres uczenia powinny zawierać informację o poziomie szczepień. Są to dane wymagane do zastosowania metody multiregresji z pseudoinwersją Moore'a-Penrose'a. W tym celu wygenerowano wykres szeregów czasowych skumulowanych szczepień dla każdego z rozpatrywanych krajów (Rys. 76. ). Na podstawie obserwacji na wykresie przyjęto datę początkową szczepień wspólną dla wszystkich krajów jako 08 października 2021 r.



Rys. 76. Wykres szeregów czasowych skumulowanej liczby szczepień dla wybranych krajów. Źródło: opracowanie własne.

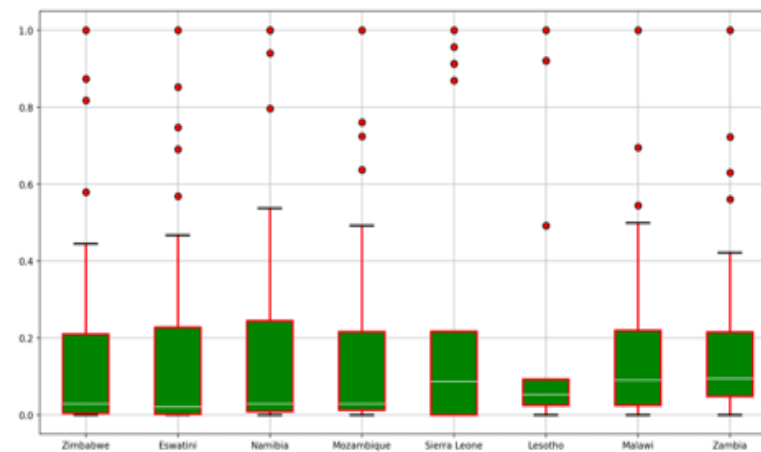
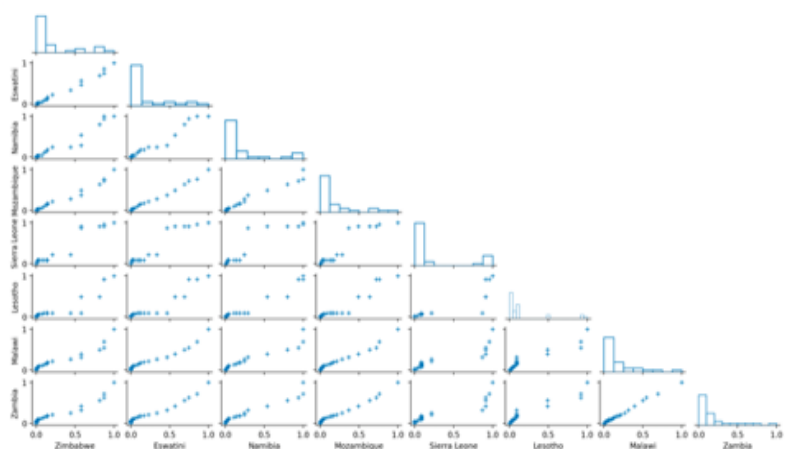
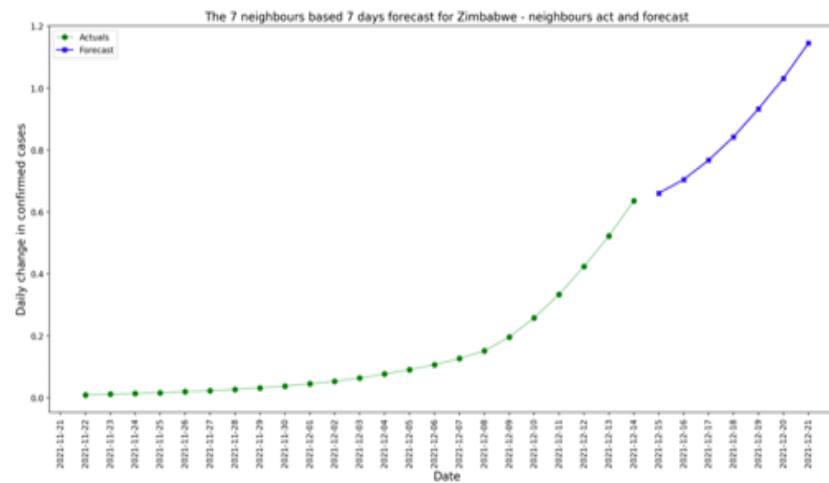
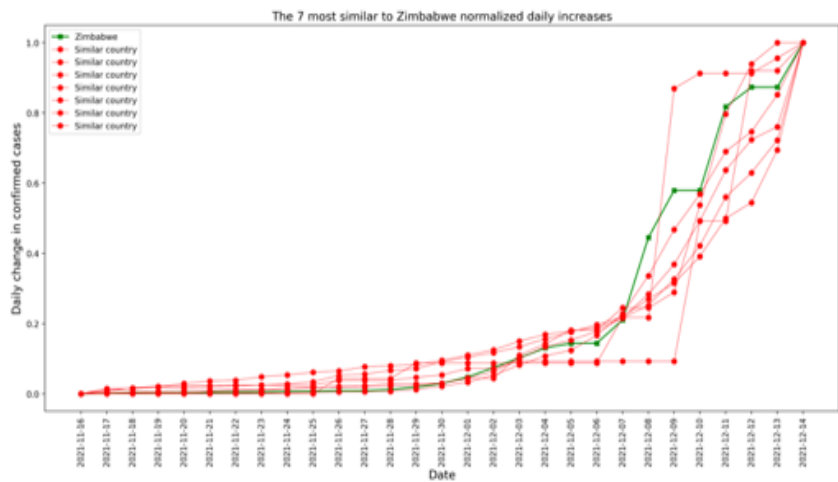
Kolejnym krokiem było wybranie zakresu dat horyzontu predykcji po 8 października 2021 r. Do celów badawczych istotne było określenie sytuacji epidemiologicznej na podstawie dynamiki zakażeń. W tym celu wygenerowano wykres szeregów czasowych skumulowanych dobowych zakażeń w wybranych krajach (Rys. 77. ). Interesującym okresem do badania okazał się przełom 2021 i 2022 roku. W krajach z Afryki była fala wznosząca zakażeń, w krajach latynoamerykańskich fala dopiero się zaczynała (Boliwia) lub miała nadejść. Kraje z Europy doświadczały wzrostów zakażeń w rozważanym zakresie dat, aczkolwiek nie tak gwałtownych jak w krajach z pozostałych kontynentów. Spośród krajów z Azji to Laos zmagał się z relatywnie dużymi wzrostami zakażeń, Kambodża utrzymywała poziom zakażeń, w Tajlandii i Wietnamie miały nadejść fale. Zdecydowano się osadzić tygodniowy horyzont predykcji w datach od 15 do 21 grudnia 2021 r., tuż przed przerwą świąteczno-noworoczną w niektórych krajach, która mogłaby mieć wpływ na raportowanie danych w oknie uczenia.



Rys. 77. Wykres szeregów czasowych skumulowanych dobowych zakażeń w wybranych krajach. Źródło: opracowanie własne.

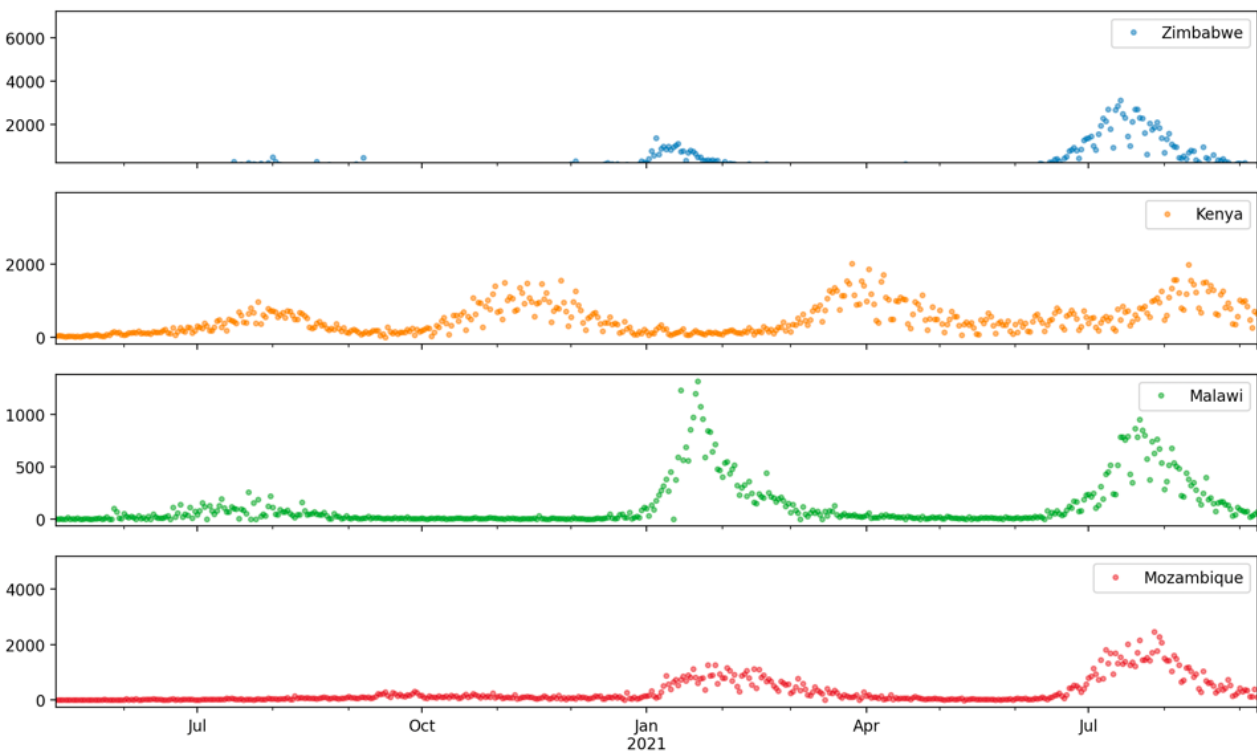
W przypadku krajów z Afryki trudno było znaleźć kraje o zbliżonych przebiegach. Analizy korelacji wzajemnych i wykresu pudełkowego pokazały liczną obecność outlierów (Rys. 78. ) oraz relatywnie niską korelację wzajemną w porównaniu do krajów z innych kontynentów.

Problemy napotkano również w przypadku badania metodą łańcuchów Markowa dla Czech, Węgier i jak się okazało większości krajów. Wykryto brak zmian dobowych zakażeń w ostatnim dniu okna czasowego, co uniemożliwiło obliczenie APE i MAPE. Skoro wzrost rzeczywistych wartości miał wartość 0, nie można było zastosować wzoru (38). Jako rozwiązanie przyjęto MAPE na podstawie predykcji dla dni, w których wykryto zmiany dobowe.



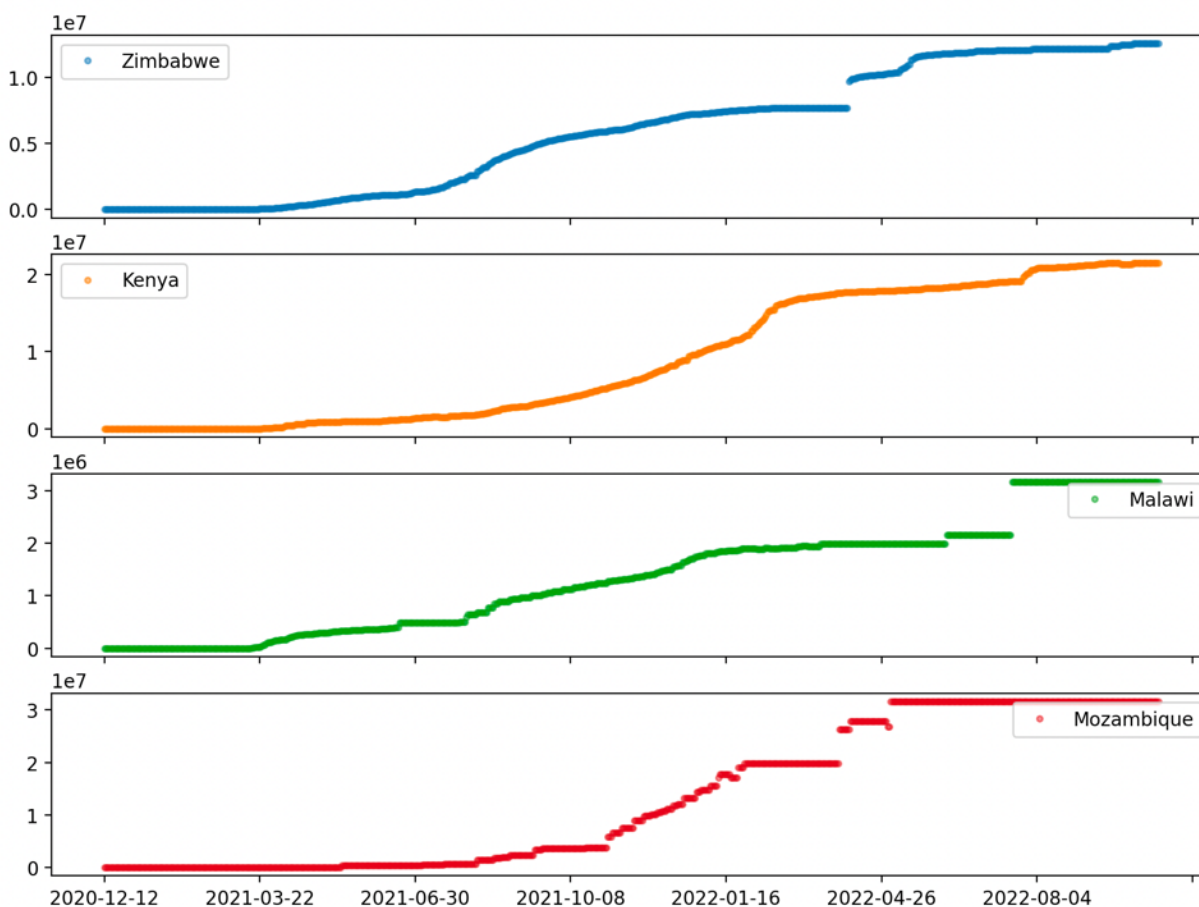
Rys. 78. Analiza graficzna danych i wyników badania metodą *n*-najbliższych sąsiadów dla kraju Zimbabwe. Źródło: opracowanie własne.

Najmniej dokładne wyniki predykcji otrzymano dla krajów afrykańskich. Kraje afrykańskie miały najmniejszą gęstość szczepień (średnia 269 w porównaniu do 1311-1400 dla pozostałych kontynentów). Przyczyną może być rzetelność raportowanych danych na poziomie dziennym lub zmiana dynamiki liczby zakażeń w tych krajach. Wygenerowany wykres szeregów czasowych przyrostów dobowych zakażeń (Rys. 79. ) potwierdził zmiany w dynamice zakażeń. Malawi i Zimbabwe (dane wejściowe i wyniki przedstawione na Rys. 78. ) wchodziło w fazę fali wznoszącej, Kenia była w fazie fali opadającej, natomiast Mozambik był tuż przed falą zakażeń.



Rys. 79. Przebiegi szeregów czasowych przyrostów dobowych zakażeń dla badanych krajów afrykańskich. Źródło: opracowanie własne.

Jakość raportowanych poddano analizie na przykładzie szczepień. Wygenerowano wykres szeregu czasowego skumulowanych szczepień dobowych dla grupy krajów afrykańskich (Rys. 80. ). W przypadku Kenii nie stwierdzono anomalii w danych, pozostałe kraje miały nagłe wzrosty, a w przypadku Mozambiku nawet pomniejszone wartości (np. w połowie stycznia 2022 r.).



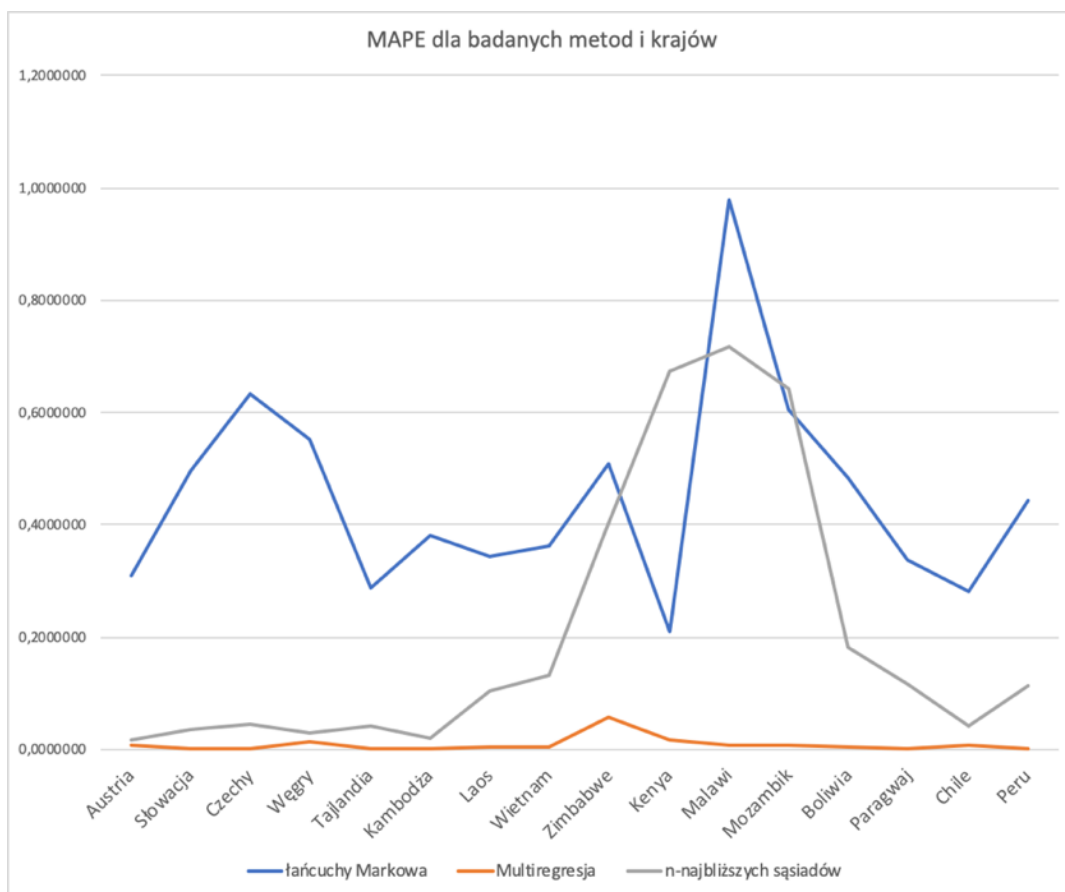
Rys. 80. Wykres szeregów czasowych skumulowanych dobowych szczepień dla badanych krajów afrykańskich. Źródło: opracowanie własne.

Wyniki MAPE uśredniono dla wszystkich metod pogrupowanych dla kontynentów. Takie podejście umożliwia agnostyczne co do metody porównanie dokładności względem położenia geograficznego badanego kraju. Najmniejsza wartość, czyli największa dokładność przypadły dla Azji (0,14), która wyprzedziła kolejno Amerykę Południową i Europę. Najmniejsza średnia dokładność predykcji – największe MAPE przypadło Afryce, która miała ~2,8 razy mniej dokładną predykcję od Azji.

Nie można jednoznacznie określić, czy położenie geograficzne wpływa na dokładność predykcji. Niższą dokładność wykrytą dla krajów afrykańskich mogły powodować czynniki związane z jakością raportowania danych, rozpoczynającymi się falami wzrostów zakażeń lub niską wartością gęstości szczepień, która co prawda nie miała wpływu na jakość predykcji w badaniu metody multiregresji z pseudoinwersją Moore'a-Penrose'a, ale może mieć wpływ na

dynamikę zakażeń. Badanie wpływu szczepień na rozwój pandemii nie jest przedmiotem niniejszej rozprawy doktorskiej.

Wyniki badania dokładności predykcji (Tab. 46. ) potwierdziły dominację metody multiregresji z pseudoinwersją Moore’a – Penrose’a (średnia MAPE dla wszystkich krajów to 0,0084), która poradziła sobie nawet z jakością danych z krajów afrykańskich. Znacznie mniej dokładne były predykcje dla krajów afrykańskich metodą n-najbliższych sąsiadów (średnie MAPE dla wszystkich krajów to 0,21), która została wykonana z parametrem  $n=7$ . Taka wartość tego parametru w poprzednich badaniach dała najbardziej dokładne prognozy. Metoda oparta na łańcuchach Markowa zawiodła dla badania większości krajów. W przypadku braku zmiany dziennych zakażeń metoda ta nie była w stanie obliczyć APE i MAPE. Metoda ta wykazała się najmniejszą dokładnością dla krajów europejskich i Malawi (Rys. 81. ) oraz relatywnie mniejszą dokładnością uśrednionego MAPE dla wszystkich krajów (0,45).



Rys. 81. Wykres ilustrujący dokładność MAPE w predykcji badanymi metodami dla wybranych krajów. Źródło: opracowanie własne.

Tab. 46. Zestawienie wyników badań na podstawie MAPE dla krajów położonych geograficznie blisko siebie. Źródło: opracowanie własne.

hp=7 15-21.12.2021 r.		Metoda oparta na łańcuchach Markowa	Multiregresja z pseudoinwersją Moore'a- Penrose'a	Metoda n- najbliższych sąsiadów	Gęstość szczepień 14.12.202 1	MAPE dla kontynentów
Europa	Austria	0,31	0,0072	0,018	1707	0,17
	Słowacja	0,50	0,0010	0,035	1016	
	Czechy	0,63	0,0011	0,044	1361	
	Węgry	0,55	0,0134	0,031	1517	
	średnia MAPE	0,50	0,0057	0,032	1400	
Azja	Tajlandia	0,29	0,0005	0,041	1405	0,14
	Kambodża	0,38	0,0002	0,022	1769	
	Laos	0,34	0,0054	0,104	801	
	Wietnam	0,36	0,0043	0,134	1365	
	średnia MAPE	0,34	0,0026	0,075	1335	
Afryka	Zimbabwe	0,51	0,0560	0,402	468	0,40
	Kenia	0,21	0,0167	0,673	152	
	Malawi	0,98	0,0064	0,719	82	
	Mozambik	0,60	0,0067	0,642	375	
	średnia MAPE	0,58	0,0214	0,609	269	
Ameryka Pd.	Boliwia	0,48	0,0060	0,181	777	0,16
	Paragwaj	0,34	0,0005	0,118	887	
	Chile	0,28	0,0082	0,042	2214	
	Peru	0,44	0,0005	0,113	1365	
	średnia MAPE	0,39	0,0038	0,114	1311	
Średnia dla metody		0,45	0,0084	0,21		

### 3.6 Badanie dokładności predykcji w zależności od fazy pandemii dla wybranego kraju.

Badanie dokładności predykcji w zależności od fazy pandemii dla wybranego kraju stanowi realizację zadań związanych z pytaniem Q4 określonym w planie prac badawczych w podrozdziale 3.1. Metody predykcji w szeregach czasowych zakażeń powinny być dokładne bez względu na dynamikę rozwoju pandemii. Poprzez dynamikę rozumiemy dobowe zmiany wartości szeregów czasowych, zwłaszcza te ekstremalne, np. poprzez początek lub koniec fali zakażeń lub szczyt fali i przejście do spadku liczby zakażeń. W tym celu zdefiniowano fazy dla wybranego kraju. Dla każdej fazy wykonano predykcję i porównano wyniki.

Dokładność predykcji dla różnych faz pandemii badano na przebiegu zakażeń w Indiach. Założono badany horyzont  $h_p=14$  dni. Do badania wykorzystano metody o potwierdzonej dotychczas wysokiej dokładności: multiregresji z pseudoinwersją Moore'a - Penrose'a oraz n-najbliższych sąsiadów ze średnią krocząca. Pominięto metodę opartą na łańcuchach Markowa ze względu na jej dokładność w tygodniowych oknach predykcji oraz niską dokładność.

Przeanalizowano wykres dobowych przyrostów zakażeń w celu określenia wyraźnych zmian w dynamice zakażeń i zidentyfikowania poszczególnych faz (Rys. 82. ). Określono następujące 4 fazy i daty rozpoczęcia 14 dniowego horyzontu predykcji:

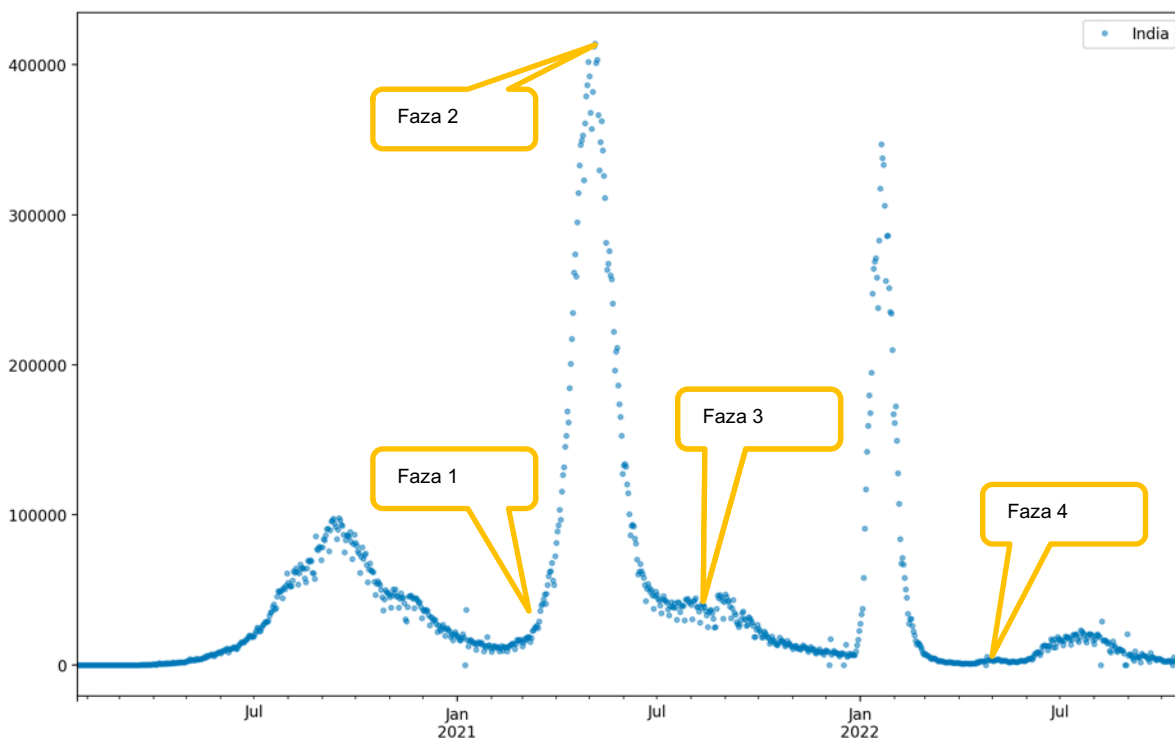
1. fali wznoszącej od 7 kwietnia 2021 r.
2. szczytu zakażeń, od 10 maja 2021 r.
3. fali opadającej, od 1 sierpnia 2021 r.
4. fazę stabilizacji, od 7 kwietnia 2022 r.

Pierwsza faza, nazwana „falą wznoszącą”, umieszczona jest w zakresie dat, w którym rozpoczęła się jedna z największych i najbardziej śmiertelnych fal krajowych pandemii COVID-19. Pierwsza faza pokrywa się z badaniem dokładności predykcji w zależności od horyzontu predykcji dla  $h_p=14$ . Zatem jedno z badań będzie dotyczyło tego samego zakresu dat horyzontu predykcji.

Druga faza to szczyt okresu zakażeń. Horyzont predykcji umieszczono tuż za wierzchołkiem, aby zakres dat okna uczenia obejmował również pojawiające się na osi czasu malejące wartości.

Trzecia faza to nazwana „fala opadająca”. Jest to zakres dat, kiedy kończy się fala zakażeń i pojawiają się pierwsze wartości szeregu czasowego z relatywnie niskimi wartościami zmian dobowych przyrostów.

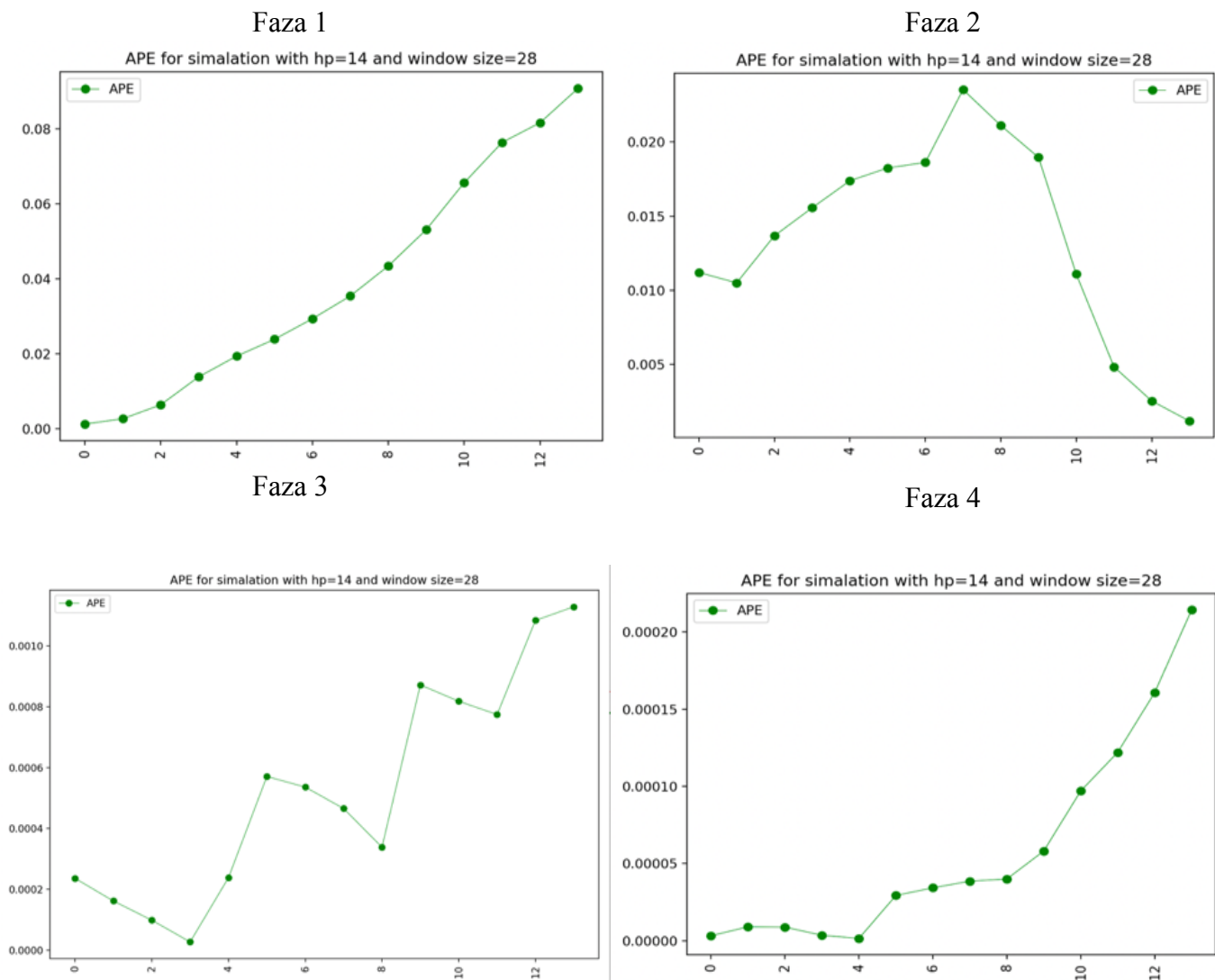
Czwarta faza to zakres dat obejmujący daty stabilizacji poziomu zakażeń, ale pojawiają się pierwsze wyższe wartości dobowych przyrostów zakażeń. Wyzwaniem dla badanych predykcji jest wykrycie potencjalnej kolejnej fali zakażeń.



Rys. 82. Wykres szeregu czasowego dobowych przyrostów zakażeń w Indiach z zaznaczonymi fazami pandemii uwzględnionymi w badaniach. Źródło: opracowanie własne.

Pierwszą badaną metodą była multiregresji z pseudoinwersją Moore’a – Penrose’a. Predykcję każdej z faz analizowano pod kontem przebiegów APE (Rys. 83. ) i wartości MAPE (Tab. 47. ) dla każdej z predykcji. W pierwszej fazie APE rośnie niemal liniowo, przyjmując największą z wartości MAPE (0,038) dla badania tą metodą. Oznacza to, że dokładność metody multiregresji z pseudoinwersją Moore’a-Penrose’a jest najniższa w przypadku początku fal pandemii. Dokładniejsze prognozy zaobserwowano w drugiej fazie, w której MAPE osiągnęło

wartość 0,013. Przebieg APE jest rosnący i tuż po minięciu połowy horyzontu predykcji maleje. Oznacza to, że metoda poradziła sobie z radykalną zmianą wartości szeregu czasowego dobowych przyrostów zakażeń. Jest to jedna z najważniejszych obserwacji w tej rozprawie, bowiem za jej pomocą można badać dokładność wprowadzanych polityk obostrzeń dysponując krótkim oknem uczenia. Dokładność MAPE znacznie polepszała się w kolejnych fazach. W trzeciej fazie – wyjścia z fali zakażeń dobowych, MAPE osiągnęło wartość 0,000928. Przebieg APE jest rosnący z okresowymi spadkami. Metoda poradziła sobie ze zmienną dynamiką dobowych przyrostów zakażeń. Czwarta faza charakteryzowała się najmniejszą dynamiką w oknie uczenia. Predykcja okazała się najdokładniejsza w rozpatrywanej metodzie i MAPE osiągnęło wartość 0,000059. Analiza przebiegu APE pokazuje stabilność do 8 dnia horyzontu predykcji, po którym następuje wzrost zbliżony do liniowego przebiegu. Spowodowane jest to nieznacznym wzrostem dobowych przyrostów zakażeń w tej fazie.

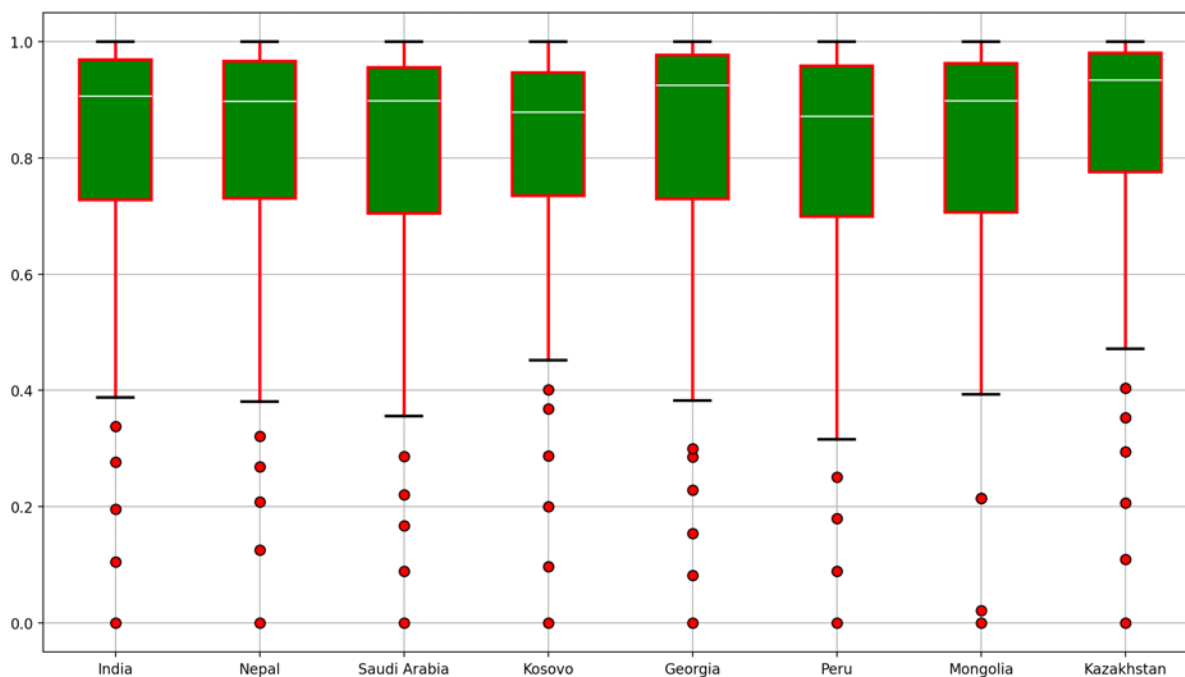


Rys. 83. Porównanie przebiegów APE dla predykcji szeregu czasowego skumulowanej liczby zakażeń w Indiach z rozróżnieniem na poszczególne fazy – metoda multiregresji z pseudoinwersją Moore’a – Penrose’a. Źródło: opracowanie własne.

W kolejno rozpatrywanej metodzie n-najbliższych sąsiadów ze średnią krocząca wykorzystano parametr  $n=7$ , ponieważ we wcześniejszych badaniach zwracał najdokładniejsze wyniki. Wykorzystano okno uczenia o długości 56 dni.

Analiza danych wejściowych - znormalizowanych wartości szeregów czasowych skumulowanych dobowych zakażeń pokazała liczną obecność outlierów, np. jak w (Rys. 84. ). Pomimo tego, udało się znaleźć kraje o zbliżonych przebiegach. W przypadku tej metody

przeanalizowano przebiegi APE i MAPE (Rys. 86. ) oraz porównani przebiegi rzeczywistych z prognozowanymi wartościami skumulowanych dobowych zakażeń (Rys. 86. i Rys. 87. ).



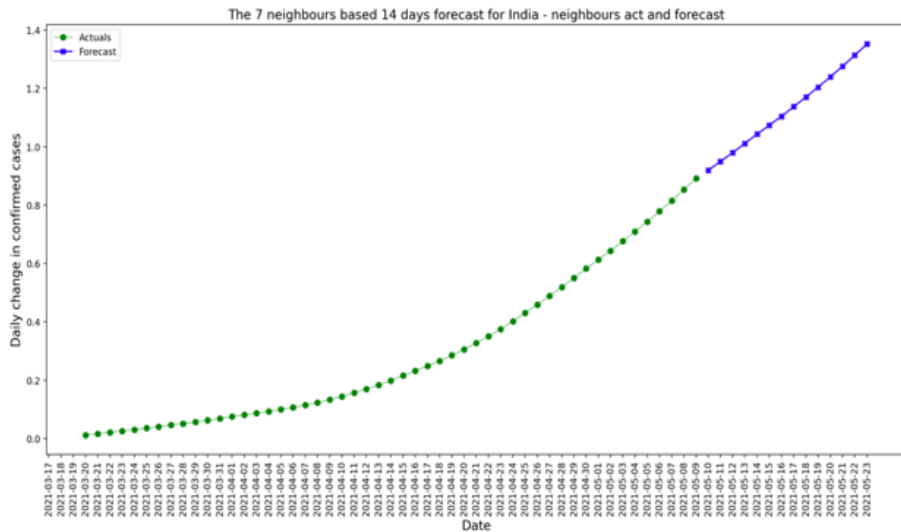
Rys. 84. Liczna obecność outlierów w analizie danych okna uczenia dla 4 fazy fragmentu szeregu czasowego skumulowanych dobowych zakażeń w Indiach. Źródło: opracowanie własne.

W pierwszej fazie APE i MAPE rosło niemal liniowo. Wartości prognozowane z każdym dniem horyzontu predykcji dzieliła większa różnica niż wartości rzeczywiste. Predykcja nie wykryła dokładnie nadchodzącej fali.

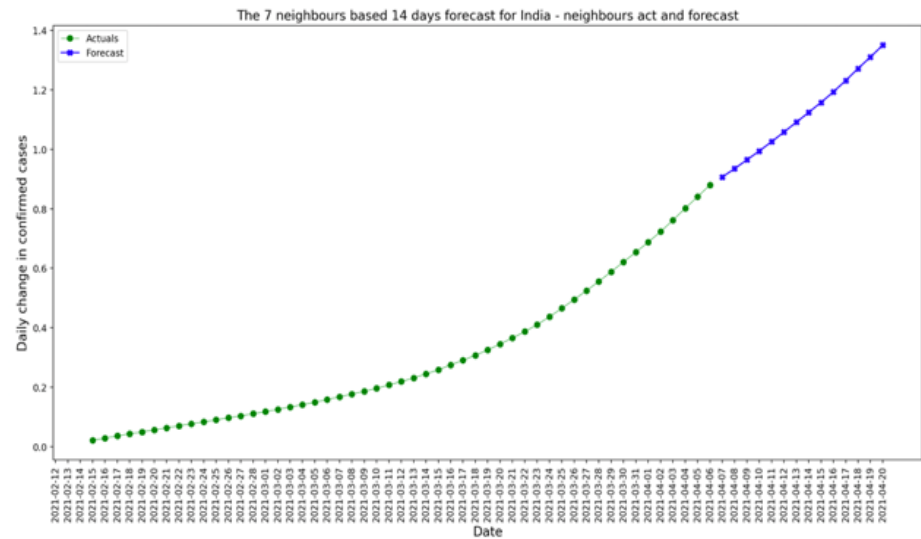
W drugiej fazie, szczytu fali zakażeń, przebiegi APE i MAPE przyjęły odwrotny kierunek niż w 1 fazie. Obie wartości malały z każdym dniem. Analiza wartości rzeczywistych i prognozowanych pokazała, że obie wartości zbiegały się w końcu horyzontu predykcji. Metoda poradziła sobie z nagłą zmianą wartości dobowych przyrostów zakażeń. Trzecia faza, obejmująca kończącą się falę zakażeń, wykazała się większą nieregularnością przebiegów APE i MAPE, które osiągnęło dokładność w horyzoncie predykcji na poziomie 0,043. Wykresy wartości rzeczywistych i prognozowanych szeregu czasowego w oknie horyzontu predykcji są niemal równoległe. Zestawienie obu szeregów na (Rys. 85. ) pokazuje niemal idealne przedłużenie

wartości prognozowanych względem rzeczywistych w 3 i 4 fazie, brak jest wyraźnego załamania na styku okna uczenia i horyzontu predykcji w porównaniu do faz 1 i 2. Najdokładniejszą predykcję osiągnięto w 4 fazie, kształtującą się na poziomie  $MAPE=0,043$ . Metoda poprawnie zareagowała na nieznaczny wzrost dobowych przyrostów zakażeń pod koniec okna uczenia się. Przebieg MAPE jest malejący w pierwszych 7 dniach horyzontu predykcji, przy czym utrzymuje przybliżony poziom do 14 dnia, natomiast APE przyjmuje przebieg zbliżony do sinusoidalnego. Krzywe wartości rzeczywistych i prognozowanych przeplatają się w horyzoncie predykcji przyjmując ostatecznie zbliżone wartości.

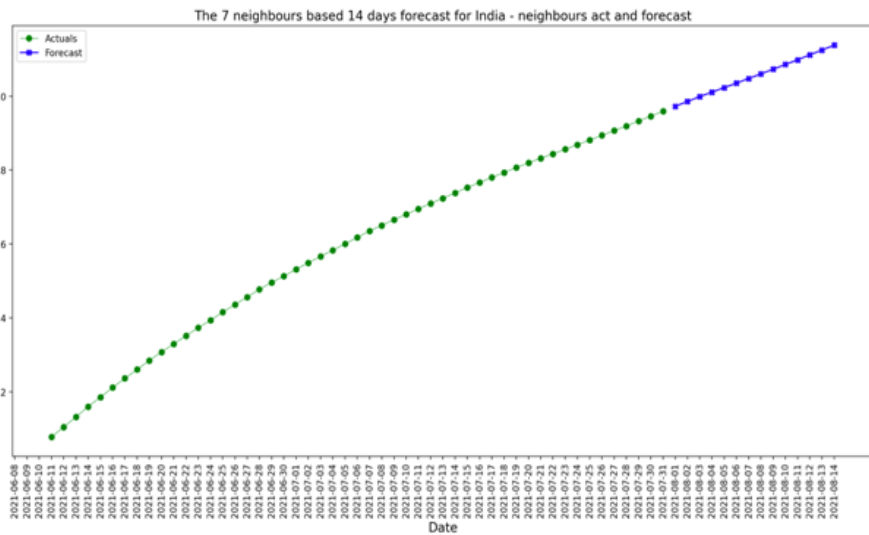
Faza 1



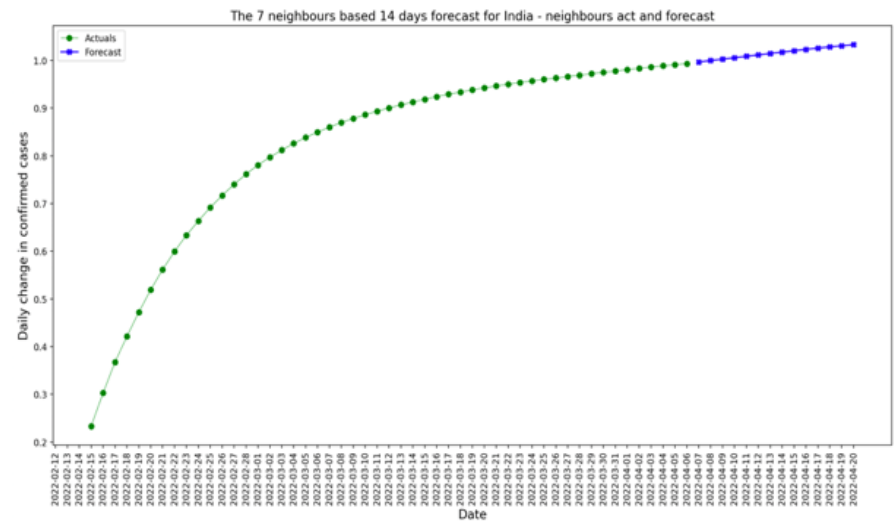
Faza 2



Faza 3

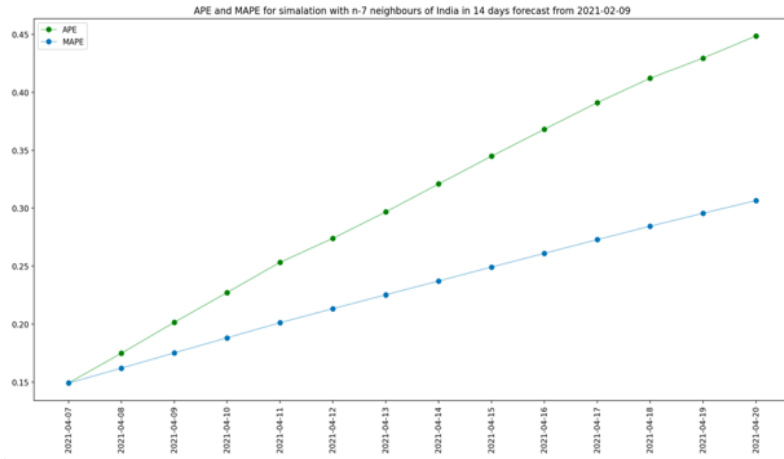


Faza 4

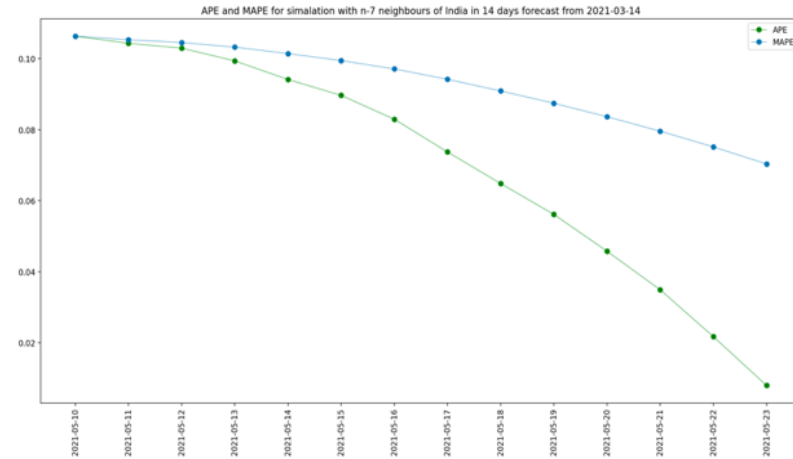


Rys. 85. Zestawienie przebiegów rzeczywistych (kolor zielony) i prognozowanych (kolor niebieski) dla predykcji szeregu czasowego skumulowanej liczby zakażeń w Indiach z rozróżnieniem na poszczególne fazy – metoda n-najbliższych sąsiadów. Źródło: opracowanie własne.

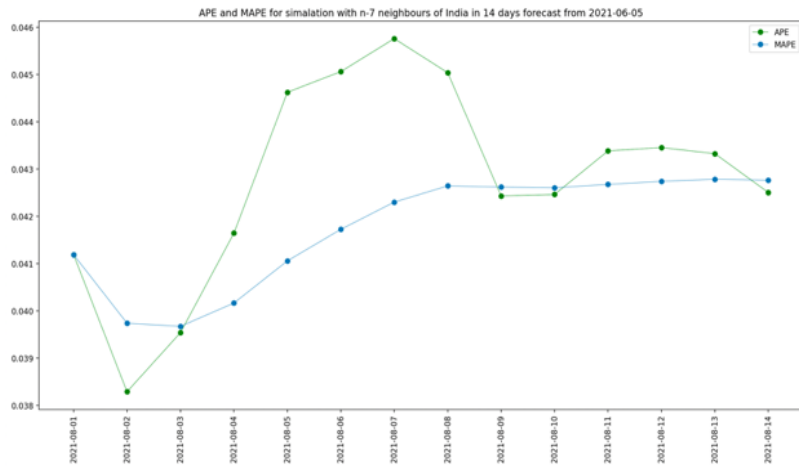
Faza 1



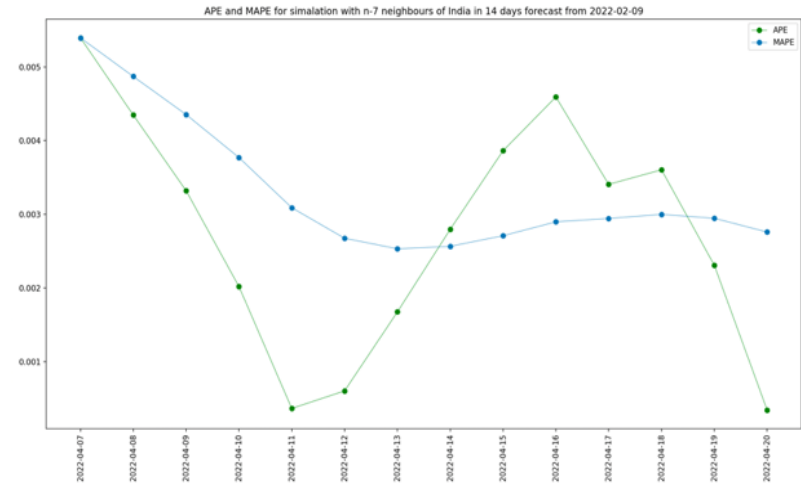
Faza 2



Faza 3

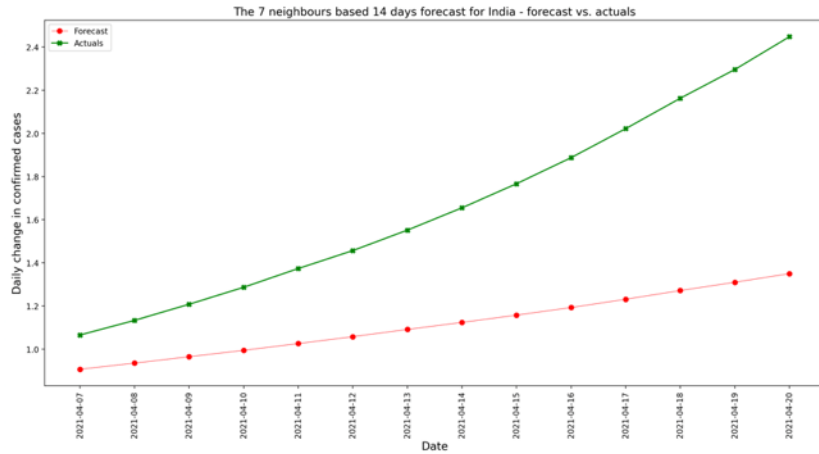


Faza 4

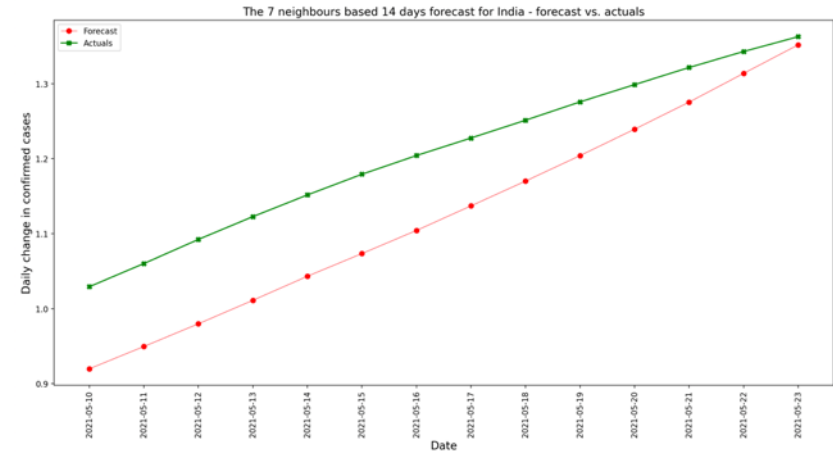


Rys. 86. Porównanie przebiegów APE i MAPE dla predykcji szeregu czasowego skumulowanej liczby zakażeń w Indiach z rozróżnieniem na poszczególne fazy – metoda n-najbliższych sąsiadów. Źródło: opracowanie własne.

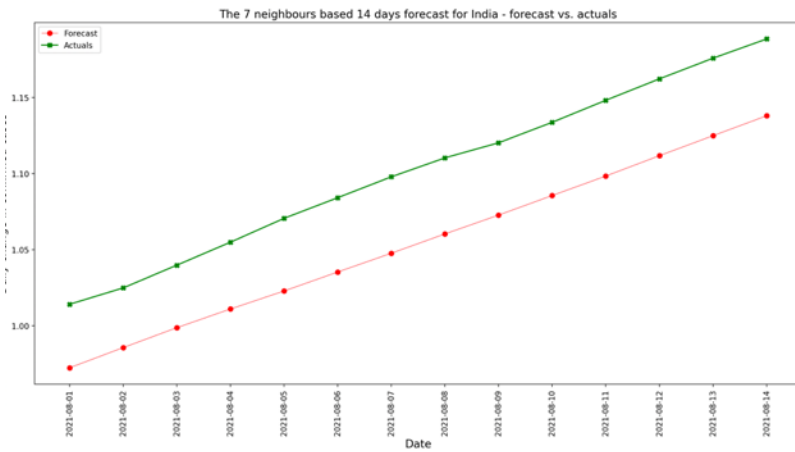
Faza 1



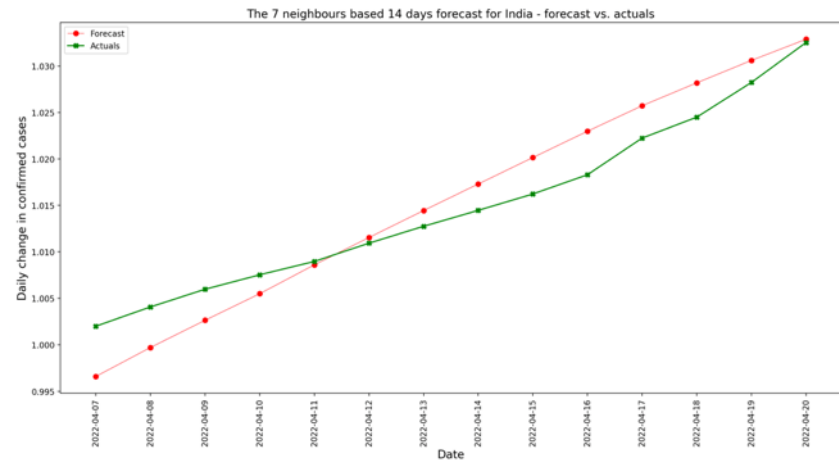
Faza 2



Faza 3



Faza 4



Rys. 87. Zestawienie przebiegów rzeczywistych (kolor zielony) i prognozowanych (kolor niebieski) dla predykcji szeregu czasowego skumulowanej liczby zakażeń w Indiach z rozróżnieniem na poszczególne fazy – metoda n-najbliższych sąsiadów. Źródło: opracowanie własne.

Zestawienie zbiorcze wyników dokładności predykcji na podstawie MAPE z rozróżnieniem na metody i fazy pandemii przedstawiono w Tab. 47.

Dokładność predykcji określona na podstawie MAPE w przypadku obu metod rosła wraz z kolejną badaną fazą. Największe trudności sprawiły dobowe przyrosty zakażeń poprzedzające falę pandemii w 1 fazie. Najlepiej obie metody poradziły sobie w przypadku końca fali pandemii oraz w czasie stabilizacji poziomu dobowych przyrostów zakażeń (3 i 4 faza). Najistotniejszą obserwacją jest dokładność predykcji w 2 fazie, kiedy kolokwialnie „wirus był w odwrocie” i liczba dobowych przyrostów zakażeń zaczynała maleć. Świadczy to o tym, że obie metody mogą znaleźć zastosowanie do prognozowania rozwoju liczby zakażeń tuż po zmianach w polityce walki z wirusem na poziomie krajowym.

*Tab. 47. Zbiorcze zestawienie wyników MAPE z predykcji szeregów czasowych skumulowanych dobowych zakażeń w poszczególnych fazach pandemii w Indiach z rozróżnieniem na metodę predykcji.*

			Multiregresja z pseudoinwersją Moore'a-Penrose'a	N=7 n-najbliższych sąsiadów	
Faza	Dynamika dobowych przyrostów zakażeń	Data rozpoczęcia okna predykcji (hp=14)	MAPE	MAPE	Data rozpoczęcia okna uczenia
1	fala wznosząca	7.04.2021	0,039	0,307	9.02.2021
2	szczyt zakażeń	10.05.2021	0,013	0,070	14.03.2021
3	fala opadająca	1.08.2021	0,00093	0,043	5.06.2021
4	faza stabilizacji	7.04.2022	0,00006	0,003	9.02.2022
Średnia			0,0133	0,1056	

Dokładność predykcji dla każdej z rozpatrywanych w badaniu metod zmierzono na poziomie uśrednionych MAPE. Metoda multiregresji z pseudoinwersją Moore'a – Penrose'a, podobnie jak w poprzednich badaniach, okazała się bardziej dokładna od metody n-najbliższych sąsiadów, tym

razem aż blisko 8 razy (pod względem MAPE). Wszystkie rozpatrywane w badaniu metody okazały się dokładne dla zdefiniowanych faz pandemii w Indiach.

## 4 Prognozowanie inflacji - metoda oparta na skorygowanym modelu regresji

Rozdział rozpoczyna się opisem przygotowania danych wraz ze wskazaniem ich źródła. Istotnym było pobranie danych z tego samego źródła, ponieważ dane powinny być gromadzone i oparte na tych samych wskaźnikach. Kolejnym krokiem było wybranie krajów. Zaproponowano kryteria określające wzorce danych oraz kraje najbardziej spełniające kryteria. Następnie określono plan badania za pomocą pytań, analogicznie jak w przypadku badania szeregów „pandemicznych”.

### 4.1 Przygotowanie danych

Dane o poziomie inflacji badanych krajów zostały pobrane ze strony OECD (<https://data.oecd.org/price/inflation-cpi.htm>). Wybrano dane wszystkich dostępnych państw oraz maksymalny zakres dat obejmujący zgromadzone dane. Pominięto dane zagregowane dla wielu państw, np. grup EU27 – obejmującej kraje Unii Europejskiej, OECD, G7 i G20. W wstępnej analizie danych wykluczono z badania kraje zawierające mniej niż 101 rekordów, ponieważ taki jest minimalny zbiór danych niezbędny do przeprowadzenia badania za pomocą zaimplementowanego algorytmu. Następnie wyodrębniono dane dla każdego kraju osobno i zapisywano je w postaci pliku MATLAB z nazwą w formacie duże CPI\_<kod\_kraju>.m w postaci wektora Y.

Badanie zostanie wykonane na grupie krajów. Przeanalizowano 100 ostatnich elementów szeregu czasowego inflacji ze zbioru danych. Analizie poddano następujące aspekty w połączeniu z metrykami:

- największa zmienność (najwyższe odchylenie standardowe).
- kraje z ekstremalnymi obserwacjami (najwyższy maksymalny wynik Z - standardowe wyniki odchylenia).

- najsilniejszy trend wzrostowy (najwyższa średnia inflacja)
- kraje z istotnym skosem (najbardziej negatywny skos)
- Największa sezonowość (różnica wariancji)

Uzyskane wyniki analizy wyłoniły czołowe 4 kraje spełniające dane kryterium.

Tab. 48. Wybór krajów do badania na podstawie kryteriów oraz lista 4 krajów o najbardziej pożądaney wartości metryk. (Źródło: opracowanie własne.)

Kryterium (Metryka)	4 kraje o najlepszych wartościach (wartość metryki)
Największa zmienność (odchylenie std)	LTU (21,38), HUN (6,77), TUR (6,71), EST (6,65)
Ekstremalne obserwacje (max Z)	DNK (3,71), CHN (3,30), NLD (3,24), BEL (3,13)
Najsilniejszy trend (średnia)	HUN (22,54), BRA (6,59), TUR (5,99), RUS (5,61)
Najbardziej negatywny skos (skosność)	SAU (0,40), MEX (0,22), ZAF (-0,01), IND (-0,27)
Największa sezonowość (różnica wariancji)	LTU (269,43), HUN (31,48), TUR (23,38), EST (22,24)

Następnie stworzono listę unikalnych krajów wyłonionych w analizie:

BEL, BRA, CHN, DNK, EST, HUN, IND, LTU, MEX, NLD, RUS, SAU, TUR, ZAF

Do grona badanych krajów dodano POL, USA oraz DEU ze względu na analizę danych krajów w innych badaniach.

Do badania metod predykcji wykorzystano również podejście GQM. Postawiono następujące pytania:

- P1: która metoda predykcji daje najbardziej przybliżone wartości rzeczywistym wyniki w jednomiesięcznym horyzoncie predykcji ( $hp = 1$ )?
- P2: jaki wpływ na dokładność predykcji ma optymalizacja parametru wygładzania  $p$ ?
- P3: jak wpływa ECM na jakość predykcji?

## 4.2 Odczyt uzyskanych wyników

Badanie przeprowadzono z wykorzystaniem zaproponowanych metod, w tym z wykorzystaniem korekcji błędu ECM oraz dodatkowo z optymalizacją w przypadku regresji wygładzania splajnem. Poniżej przedstawiono metryki wraz z wzorami, którymi mierzono dokładność prognoz. Badanie wykonano na 100 ostatnich obserwacjach dla każdego badanego kraju. Prognozowano 75 elementów szeregu czasowego i porównano je z wartościami rzeczywistymi. Pierwsze 25 elementów stanowiło okno uczenia o liczbie elementów odpowiadającej maksymalnej wartości parametru *kwst*.

W analizie posłużono się kodami krajów. Ułatwiło to czytelność wyników oraz implementację kodu. Państwa o kodach ISO 3166-1 alpha-3: BEL, BRA, CHN, DEU, DNK, EST, HUN, IND, LTU, MEX, NLD, POL, RUS, SAU, TUR, USA i ZAF to odpowiednio Belgia, Brazylia, Chiny, Niemcy, Dania, Estonia, Węgry, Indie, Litwa, Meksyk, Holandia, Polska, Rosja, Arabia Saudyjska, Turcja i RPA. Skrypt dokonujący predykcji uruchomiono iteracyjnie dla danych z powyższej grupy państw.

Każdorazowe uruchomienie skryptu dla danego kodu kraju zwraca tabelę wyników w postaci pliku CSV o nazwie zawierającej konkatencję znacznika czasowego i kodu kraju.

Dla każdego badanego kraju wykonano predykcje następującymi metodami:

1. Regresją wielomianową z optymalizacją parametrów *kwst* i *n*.
2. Korekcji błędu ECM bazując na prognozowanych wartościach uzyskanych w metodzie z pkt. 1
3. Regresję wygładzania splajnem z optymalizacją parametru *kwst*.
4. Regresję wygładzania splajnem z optymalizacją parametrów *kwst* i *p*.
5. Korekcji błędu ECM bazując na prognozowanych wartościach uzyskanych w metodzie z pkt. 4

Wyniki symulacji w postaci wartości rzeczywistych i prognozowanych wybranymi metodami zebrano w plikach CSV dla każdego z badanych krajów z osobna.

Nazwy metod predykcji zostały uwzględnione w kodzie i na wykresach w następujący sposób:

- Polynomial\_Prediction – predykcja regresją wielomianem

- Polynomial\_Prediction\_ECM - predykcja regresją wielomianem z korekcją ECM
- Spline\_Opt\_Prediction – predykcja regresją wygładzania splajnem z optymalizacją.
- Spline\_Opt\_Prediction\_ECM - regresją wygładzania splajnem z optymalizacją i korektą ECM.
- Spline\_Prediction - regresją wygładzania splajnem z zastosowaniem domyślnej metody wyboru parametru wygładzania zgodnie ze wzorem (57-1).

Wyniki oceniono pod względem metryk:

- MAPE (Mean Absolute Percentage Error) – w przypadku inflacji będzie liczone w następujący sposób:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{|Y_i|} \times 100\% \quad (58)$$

gdzie:

$n$  – liczba obserwacji,

$Y_i$  – wartość rzeczywista,

$\hat{Y}_i$  – wartość prognozowana.

- MSE (Mean Squared Error):

$$MSE_{hp} = \frac{1}{n} \sum_{i=1}^n (Y_{i+hp} - \hat{Y}_{i+hp})^2 \quad (59)$$

gdzie:

$Y_{i+hp}$  – rzeczywiste wartości w horyzoncie predykcji  $hp$ ,

$\hat{Y}_{i+hp}$  – przewidywane wartości.

- RMSE (Root Mean Squared Error):

$$RMSE_{hp} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{i+hp} - \hat{Y}_{i+hp})^2} \quad (60)$$

gdzie:

$Y_{i+hp}$  – rzeczywiste wartości w horyzoncie predykcji  $hp$ ,

$\hat{Y}_{i+hp}$  – przewidywane wartości.

- MAE (Mean Absolute Error):

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_{i+hp} - \hat{Y}_{i+hp}| \quad (61)$$

gdzie:

$Y_{i+hp}$  – rzeczywiste wartości w horyzoncie predykcji  $hp$ ,

$\hat{Y}_{i+hp}$  – przewidywane wartości.

- $R^2$  (R-squared):

$$R_{hp}^2 = 1 - \frac{\sum_{i=1}^n (Y_{i+hp} - \hat{Y}_{i+hp})^2}{\sum_{i=1}^n (Y_{i+hp} - \bar{Y}_{hp})^2} \quad (62)$$

gdzie:

gdzie:

$Y_{i+hp}$  – rzeczywiste wartości w horyzoncie predykcji  $hp$ ,

$\hat{Y}_{i+hp}$  – przewidywane wartości.

$\bar{Y}_{hp}$  – średnia arytmetyczna rzeczywistych wartości w horyzoncie  $hp$ .

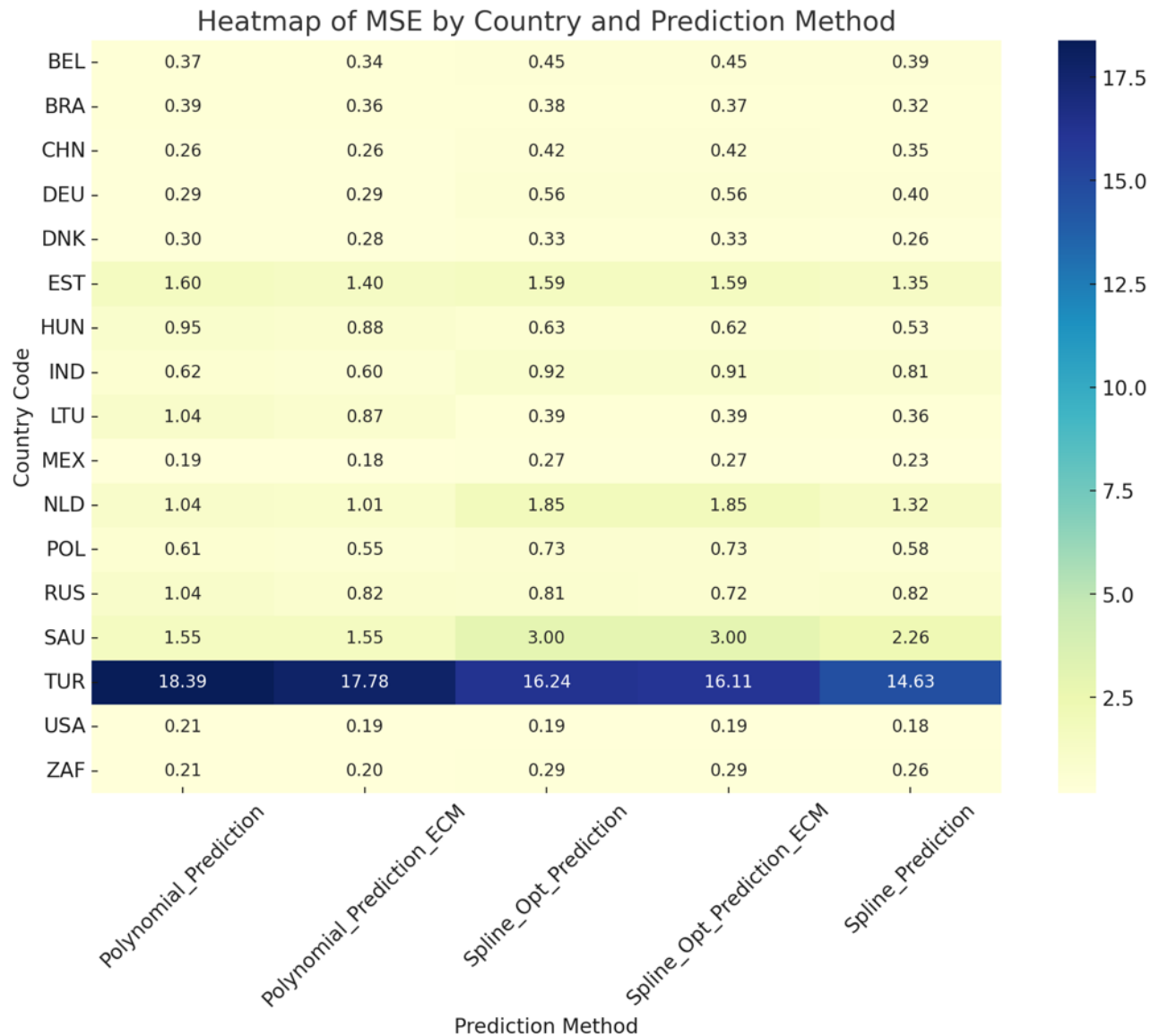
$$\bar{Y}_{hp} = \frac{1}{n} \sum_{i=1}^n Y_{i+hp} \quad (63)$$

### 4.3 Pomiar dokładności metody

Wyniki badania przedstawiono w postaci zbiorczej tabeli zawierającej metryki dla danego kraju i metody predykcji oraz zapisano w jednym pliku. Poniżej przedstawiono wartości MSE i RMSE w postaci map ciepła. Porównano wpływ optymalizacji w przypadku regresji z wygładzaniem splajmem na dokładność predykcji względem domyślnych wartości parametru wygładzania. Przeanalizowano wpływ korekcji ECM na dokładność predykcji oraz zobranowano na wykresach. Ekstremalne wyniki metryk zinterpretowano oraz dla wybranych przypadków (np. stabilnej inflacji w USA oraz ekstremalnych wartości dla Rosji i Turcji) przedstawiono na wykresach.

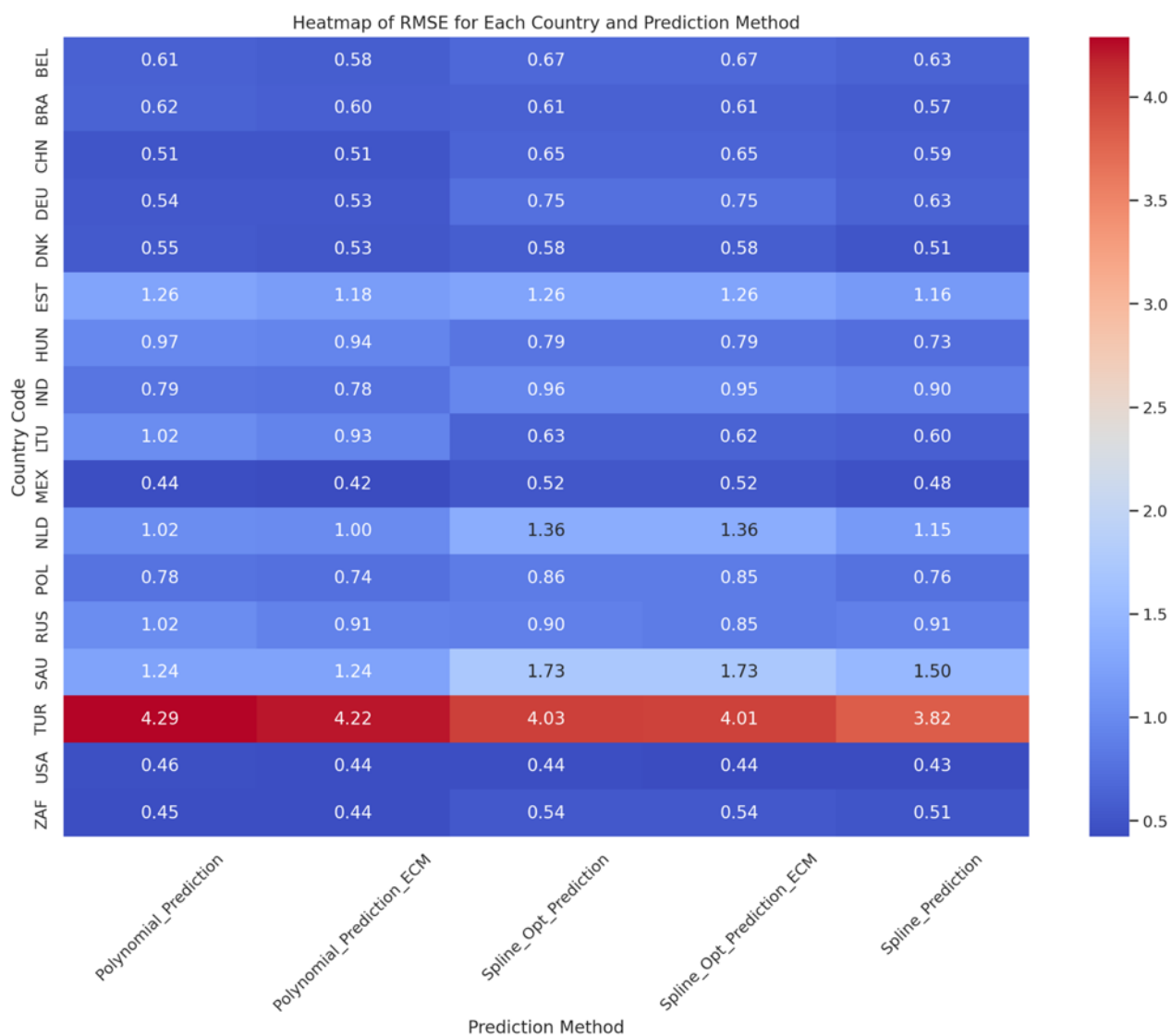
Analiza MSE dla różnych metod prognozowania (Tab. 49. ) wykazała, że predykcja regresją wygładzającą splajmem jest najbardziej dokładną metodą z najniższym średnim MSE (1,47), sugerując przewagę technik splajnow nad metodami wielomianowymi ("Polynomial\_Prediction" i "Polynomial\_Prediction\_ECM" z MSE około 1,71 i 1,62). Optymalizacja regresji wygładzania splajmem nie przyniosła znaczącej poprawy, co może wskazywać na efektywność podstawowej metody splajnow. Zastosowanie korekcji ECM w metodach wielomianowych minimalnie obniżyło MSE, co wskazuje na ograniczoną dokładność ECM w tym kontekście.

Tab. 49. Zbiorcze zestawienie wartości metryki predykcji MSE dla wybranych krajów w postaci mapy ciepła. Źródło: opracowanie własne.



Na mapie ciepła (Tab. 50. ) przedstawiono wartości RMSE dla różnych metod predykcji w poszczególnych krajach. Niższe wartości RMSE są pożądane, ponieważ wskazują na większą dokładność modelu. Modele predycyjne dla Niderlandów (NLD) i Stanów Zjednoczonych (USA) wykazywały stosunkowo niskie wartości RMSE we wszystkich metodach, co sugeruje, że metody te były wyjątkowo dokładne w przewidywaniach dla tych krajów.

Tab. 50. Zbiorcze zestawienie wartości metryki predykcji RMSE dla wybranych krajów w postaci mapy ciepła. Źródło: opracowanie własne.



Przeciwnie, w Turcji (TUR) obserwowano znacznie wyższe wartości RMSE dla wszystkich metod, co wskazuje na to, że żadna z testowanych metod nie była szczególnie dokładna w przewidywaniu wyników dla tego kraju, co może być wynikiem unikalnych wyzwań związanych z danymi. W przypadku Indii (IND) i Rosji (RUS), RMSE był zbliżony między różnymi metodami predykcyjnymi, co wskazuje na to, że charakter danych dla tych krajów był podobnie przewidywalny dla wszystkich stosowanych metod.

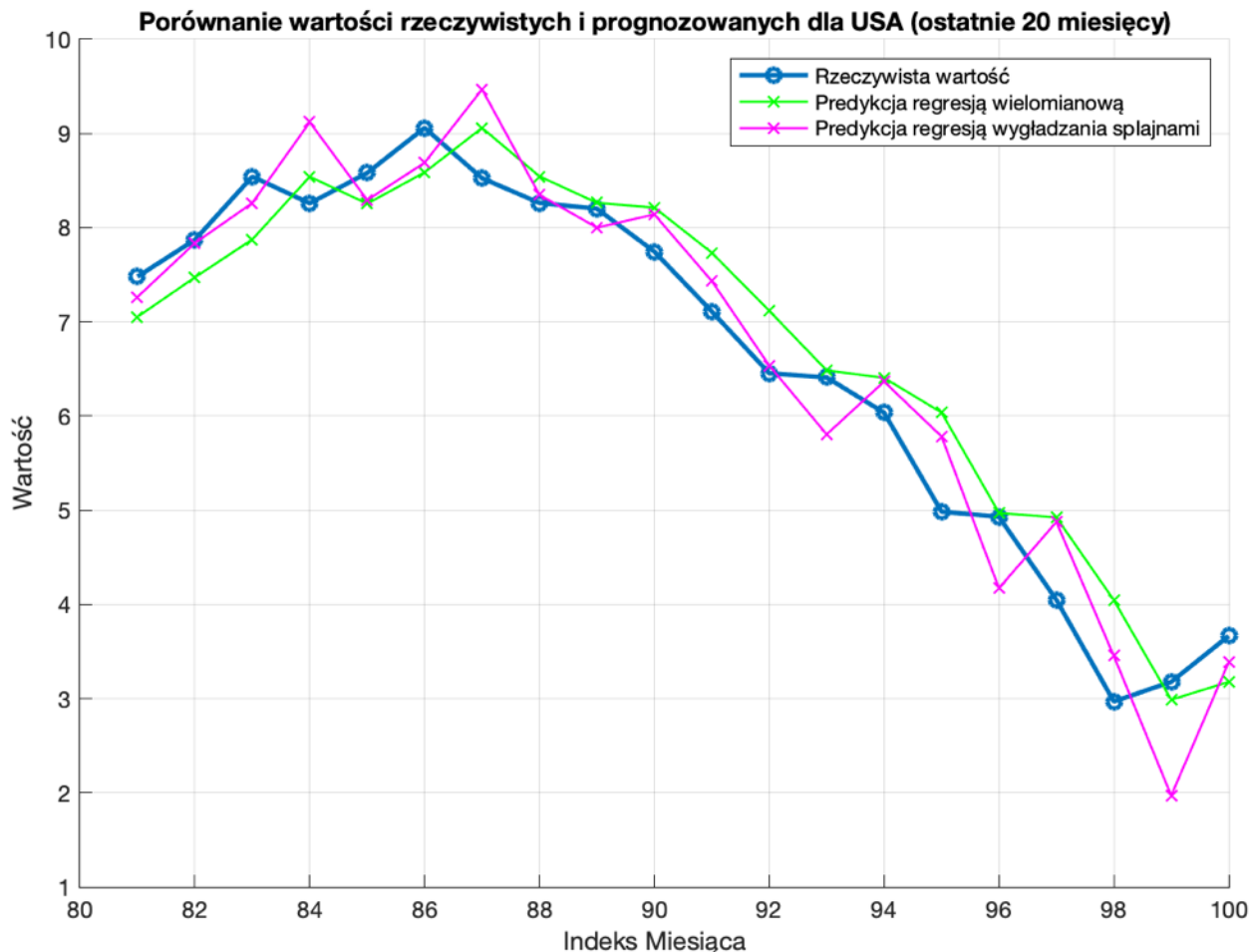
Tab. 51. Zbiorcze porównanie średnich wartości metryk prognozowania zastosowanymi metodami. (Źródło: opracowanie własne.)

Metoda predykcji	MSE	RMSE	MAE	MAPE	R <sup>2</sup>
Polynomial_Prediction	1,71	0,98	0,61	16,49	0,91
Polynomial_Prediction_ECM	1,62	0,94	0,61	17,85	0,92
Spline_Opt_Prediction	1,71	1,02	0,65	20,63	0,88
Spline_Opt_Prediction_ECM	1,69	1,01	0,65	20,71	0,88
Spline_Prediction	1,47	0,93	0,60	18,50	0,90

Zbiorcze zestawienie rozpatrywanych metryk dla wszystkich metod przedstawiono w Tab. 51.

Regresja wielomianowa z ECM i regresja wygładzająca splajnem są najbardziej dokładnymi metodami. Ta druga ma najniższy średni MSE o 0,24 punktu niższe niż regresja wielomianowa i zastosowanie optymalizacji w tej metodzie, co przekłada się na około 14% poprawę. Jednakże, regresja wygładzająca splajnem ma średni R<sup>2</sup> równy 0,90, co oznacza gorsze dopasowanie do danych o 2,2% niż regresja wielomianowa z ECM.

Najniższą wartość MSE dla najbardziej dokładnej z metod osiągnięto dla predykcji inflacji w USA. Na wykresie przedstawiono porównanie wartości rzeczywistych z prognozowanymi najdokładniejszymi metodami.

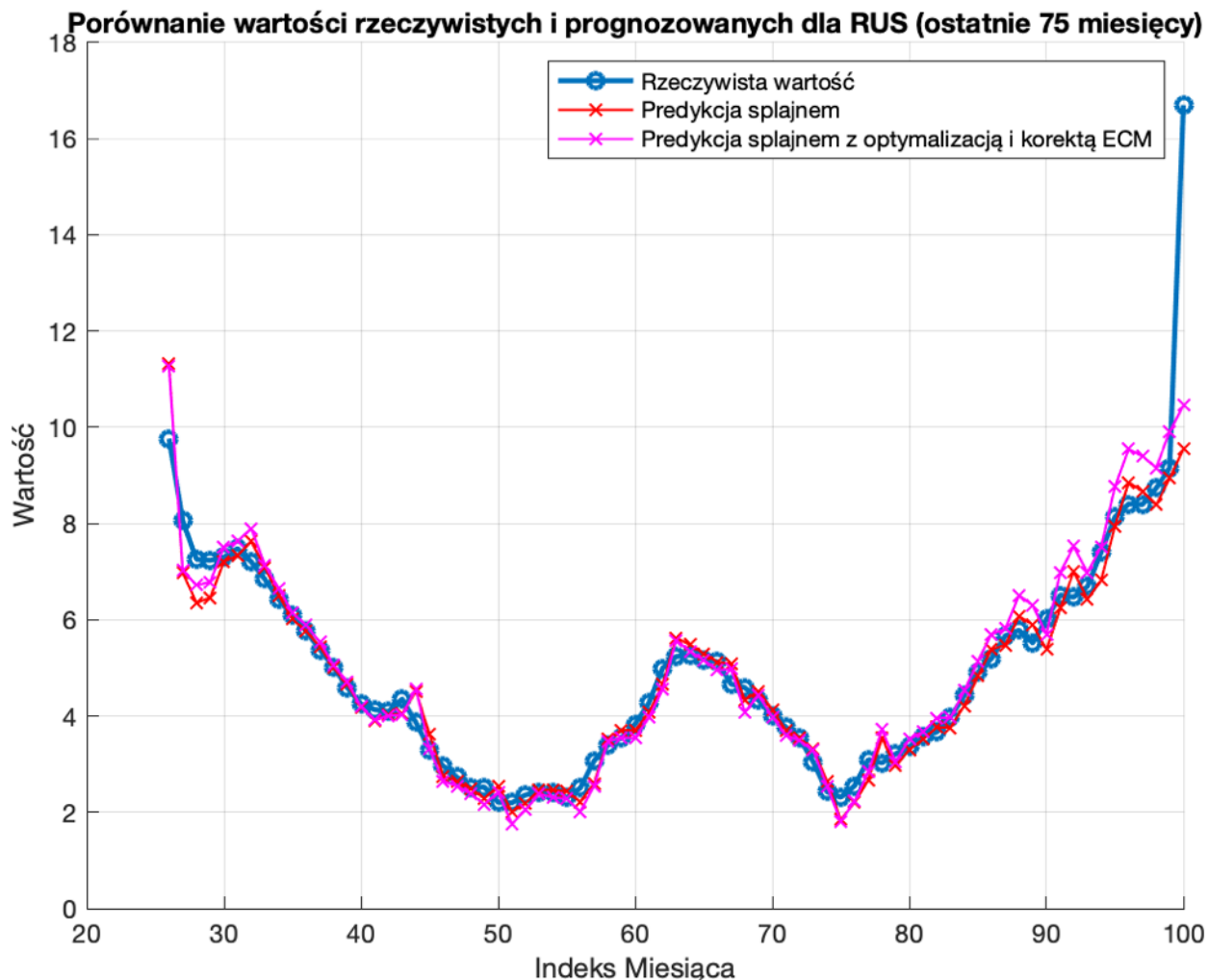


Rys. 88. Porównanie 20 ostatnich wartości rzeczywistych i prognozowanych (metody regresji wielomianowej (kolor zielony) i wygładzania splajnem (kolor magenta) dla USA. Źródło: opracowanie własne.

### Zastosowanie optymalizacji parametru $p$ .

Po optymalizacji regresji wygładzającej splajnem, średni MSE wzrósł o około 16%, a średni RMSE wzrósł o około 10%. Zastosowanie korekty ECM na regresji wygładzającej splajnem z optymalizacją spowodowało wzrost średniego MSE o około 15%, natomiast RMSE wzrósł o około 8%. Wartości  $R^2$  obniżyły się o około 2,2% w obu przypadkach optymalizowanych, co wskazuje na nieznaczne pogorszenie zdolności modelu do wyjaśniania wariacji danych.

Dla Rosji (Rys. 89. ) optymalizacja regresji wygładzającej splajnem okazała się lepsza niż prognoza bez optymalizacji.



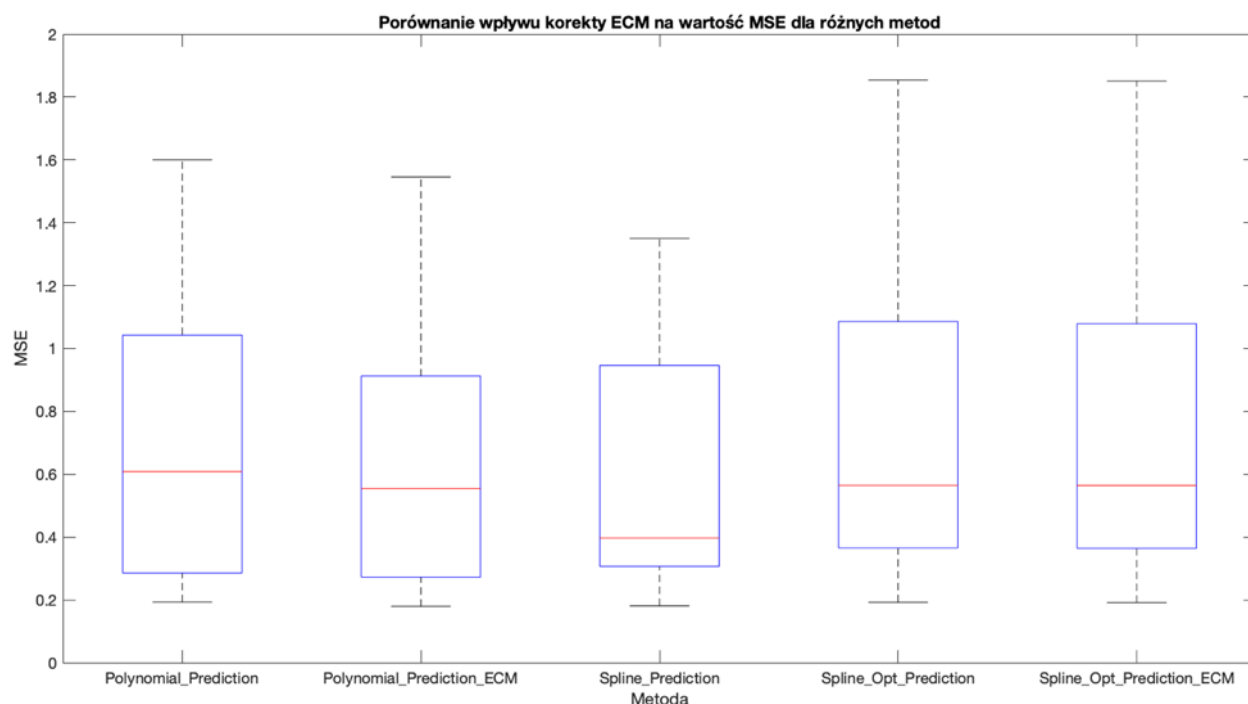
Rys. 89. Porównanie 75 ostatnich wartości rzeczywistych i prognozowanych (metody regresji wygładzania splajnem bez optymalizacji (kolor czerwony) i wygładzania splajnem z korektą ECM (kolor magenta) dla Rosji. Źródło: opracowanie własne.

W przypadku Rosji wszystkie kluczowe wskaźniki błędów (MSE, RMSE, MAE, MAPE) były niższe dla prognozy z optymalizacją w porównaniu z prognozą bez optymalizacji. Graficzna analiza przebiegów na wykresie wskazuje, że przebiegi wartości prognozowanych z optymalizacją i bez są zbliżone, nawet z wyraźną przewagą predykcji bez optymalizacji w ostatnich 15 prognozach. Niższe wartości metryk bez optymalizacji najprawdopodobniej spowodowane są specyfiką danych rzeczywistych, które wyróżniały się silnym trendem wzrostowym w końcowej części szeregu czasowego.

#### Zastosowanie korekty ECM.

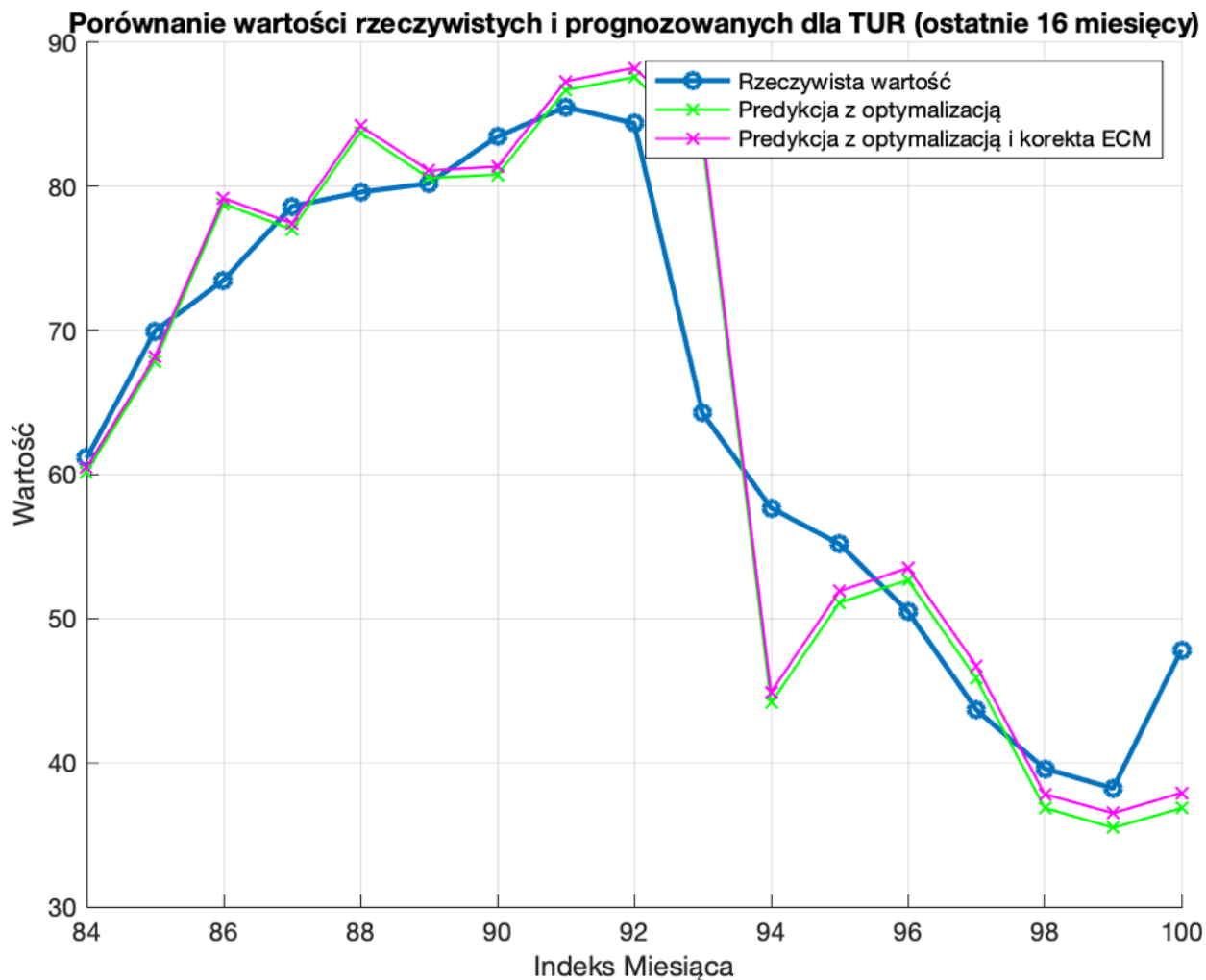
Po zastosowaniu korekty błędu ECM w regresji wielomianowej: średni MSE zmniejszył się o około 5,3%, a średni RMSE zmniejszył się o około 4,1%. Podobnie, dla metody regresji wygładzającej splajnem, stosowanie ECM zmniejszyło średni MSE o około 1,2% i RMSE o około 1%. Wartości  $R^2$  w obu przypadkach zastosowania ECM lekko wzrosły, co wskazuje na nieco lepsze dopasowanie modelu.

Wpływ korekty ECM (i jednocześnie optymalizacji) na MSE wszystkich symulacji dla metod predykcji przedstawiono na wykresie pudełkowym (102).



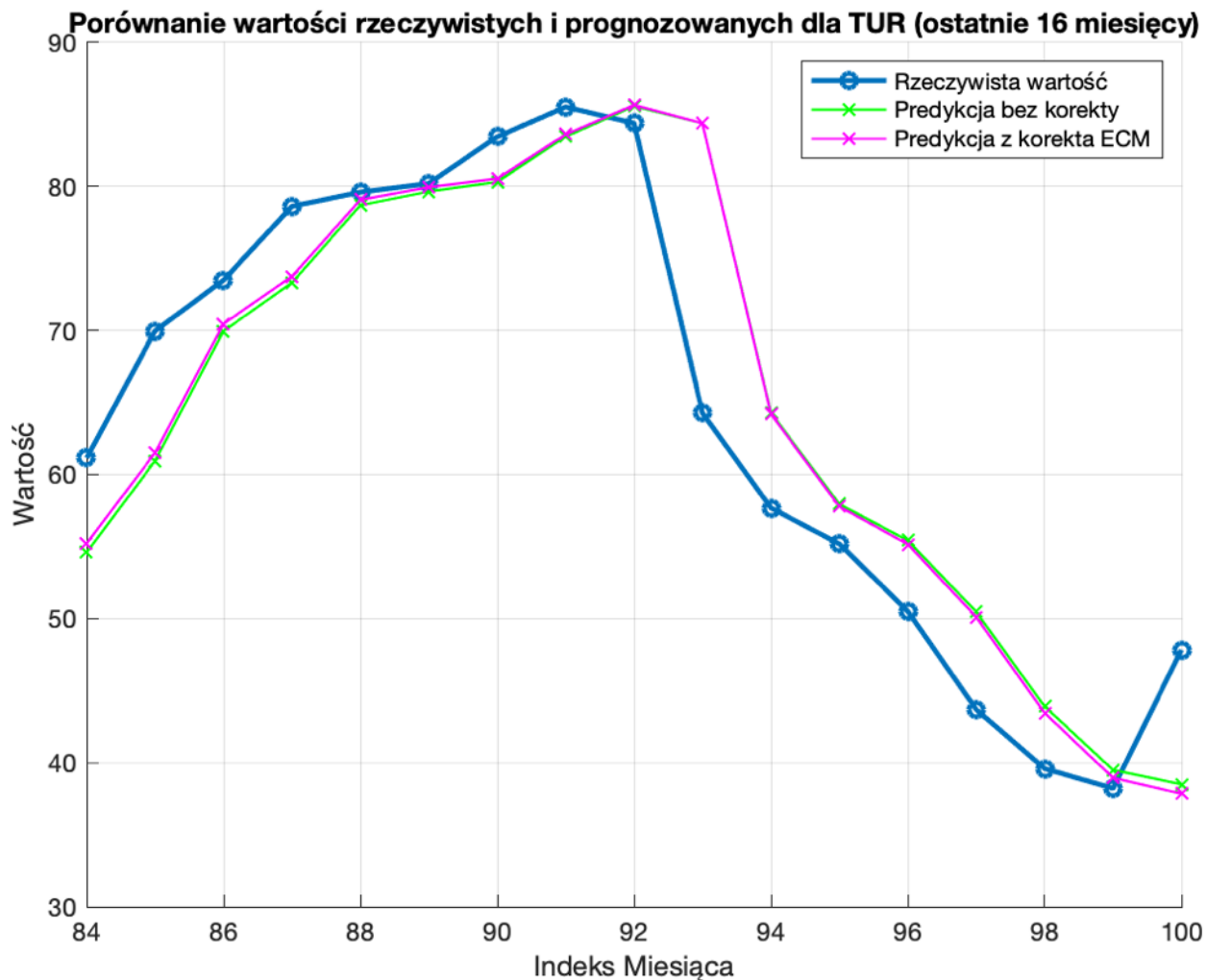
Rys. 90. Porównanie wpływu optymalizacji i korekty ECM na MSE dla zastosowanych metod predykcji na całym zbiorze danych wejściowych. Źródło: opracowanie własne.

Największa różnica MSE (0,13) pomiędzy predykcją z ECM i bez jej zastosowania w przypadku metody regresji z wygładzania splajnem z optymalizacją była dla Turcji (Rys. 91. ).



Rys. 91. Porównanie 16 ostatnich wartości rzeczywistych i prognozowanych (metody regresji wygładzania splajnem z optymalizacją (kolor zielony) i optymalizacji połączonej z korektą ECM (kolor magenta) dla Turcji. Źródło: opracowanie własne.

W przypadku regresji wielomianem największy wpływ miała korekta ECM również w przypadku Turcji (Rys. 92. ). Różnica w MSE dla tej metody predykcyjnej w Turcji wynosiła około 0,613.



Rys. 92. Porównanie 16 ostatnich wartości rzeczywistych i prognozowanych (metody regresji wielomianem bez korekty (kolor zielony) i z korektą ECM (kolor magenta) dla Turcji. Źródło: opracowanie własne.

Wysoka zmienność i unikatowe cechy danych wejściowych dla Turcji mogły sprawić, że prognozowanie było wyzwaniem, gdyż zastosowane metody predykcyjne mogą mieć trudności z adaptacją do tak specyficznych warunków. Różnorodność i nieregularność wzorców w danych mogły dodatkowo zwiększać ryzyko błędów w prognozach.

## 5 Konkluzje

Rozdział ten skupia się na konkluzjach i dyskusji wyników badań dotyczących metod predykcji w szeregach czasowych, związanych z pandemią wywołaną przez wirus SARS-CoV-2 oraz zjawiskiem inflacji. Wyniki badań potwierdzają, że zastosowane metody są efektywne w modelowaniu predykcji w szeregach czasowych wybranych procesów złożonych, osiągając satysfakcjonującą dokładność predykcji. Rozdział ten podkreśla także znaczenie społeczne i gospodarcze prognozowania liczby zakażeń oraz inflacji.

W końcowej części rozdziału zaproponowano dalsze kierunki badań. Podkreślono uniwersalność metod oraz możliwość implementacji, np. jako usługi sieciowe. Wskazano też możliwości optymalizacji proponowanych metod.

### 5.1 Dyskusja wyników badań

W niniejszym podrozdziale przedstawiono konkluzje wynikające z badań nad metodami predykcji w szeregach czasowych dotyczących pandemii wywołanej przez wirus SARS-CoV-2 oraz zjawiska inflacji. Oba te obszary charakteryzują się znaczącą złożonością, co wymagało zastosowania zaawansowanych podejść analitycznych. W pracy zaprezentowano metody, które adekwatnie radzą sobie z tą złożonością, demonstrując ich skuteczność w przewidywaniu przyszłych trendów w danych szeregach czasowych.

Analiza pandemii COVID-19 wymagała zdobycia wiedzy z jej zakresu, ze względu na dynamiczny i nieprzewidywalny charakter rozprzestrzeniania się wirusa. Z kolei zjawisko inflacji, ze względu na jego ewolucyjny charakter i zależność od wielu czynników ekonomicznych, wymagało opracowania adaptacyjnych modeli predykcyjnych. Użyte metody okazały się być skuteczne w modelowaniu tych złożonych systemów. Mimo wyzwań związanych ze złożonością tych procesów, w badaniach osiągnięto satysfakcjonującą dokładność predykcji w różnych horyzontach czasowych. Wyniki te potwierdzają, że zastosowane metody są odpowiednie do modelowania i prognozowania w warunkach wysokiej niepewności i dynamicznych zmian, stanowiąc, zdaniem autora, cenny wkład w dziedzinę analizy szeregów czasowych.

Pandemia wywołana wirusem SARS-CoV-2 jest istotnym problemem społecznym, gospodarczym oraz politycznym o zasięgu globalnym. Wprowadzanie obostrzeń, czasami

asekuracyjnie bez żadnych racjonalnych przesłanek, powoduje liczne napięcia społeczne często niosące za sobą konsekwencje polityczne. Najbardziej istotny jest jednak wymiar humanitarny pandemii. Miliony osób straciło życie, znacznie więcej będzie borykać się ze skutkami zakażenia wirusem. Pierwsze miesiące pandemii pokazały brak narzędzi predykcyjnych wspomagających decyzje rządzących związane z przeciwdziałaniem zakażeniom oraz przygotowaniem na nadchodzące fale zakażeń. Dokładne metody predykcyjne umożliwiłyby optymalne zarządzanie zasobami. Szpitale znałyby zapotrzebowanie na łóżka szpitalne, tlen, personel medyczny, itp. Wprowadzane polityki obostrzeń mogłyby być szybko weryfikowane pod względem dokładności. W tym celu stworzono nowe lub rozbudowane liczne modele i metody statystyczne.

Przedstawiono powszechnie stosowane i opisane w literaturze modele i metody predykcyjne stosowane w prognozowaniu rozprzestrzenia się wirusa SARS-CoV-2. Spośród istniejących modeli przeanalizowano metody bazujące na modelach epidemiologicznych (SIR i pochodne), modele matematyczne (model ABM), ekonometryczne (AR, MA, ARIMA) oraz popularne modele Deep Learning.

Powszechnie stosowana metoda SIR okazała się niedokładna w początkowej fazie pandemii, bowiem nie było spójnego i wiarygodnego raportowania danych (przypadek Wuhan). Głównym problemem związanym z epidemiologicznym podejściem okazał się brak znajomości specyfiki transmisji i biologii wirusa oraz wpływu na człowieka. Model SIR był opracowywany na potrzeby chorób zakaźnych dobrze wówczas znanych człowiekowi, np. wścieklizny, dla której śmiertelność wynosi praktycznie 100%, nie licząc przeżywalności na poziomie kilku osób począwszy od 1970 r. (Sadkowska-Todys (2006)). Przebieg choroby COVID-19 wywoływanej przez wirus SARS-CoV-2 jest bardziej złożony, występują np. stany ozdrowienia, wtórne zakażenia i ozdrowienia, itp. W miarę poznawania wirusa wykształcono rozbudowane modele „pochodne” od SIR, jak SEIR i inne, które są obecnie rozwijane.

Modele matematyczne ABM polegają na założeniach związanych z kontaktami ludzkimi. O ile można monitorować mobilność ludzi danej populacji, o tyle nie wiadomo jak i czy te osoby miały kontakt, czy doszło do transmisji. Jest to czysto teoretyczny model działający na przyjętych założeniach i empirycznych parametrach korygujących, mający zastosowanie w małej populacji i w początkowej fazie choroby zakaźnej.

Ciekawym modelem okazał się ekonometryczny – ARIMA. Autor poza niniejszą rozprawą wykonał również badania predykcji metodą ARIMA, wyniki pokrywały się z dotychczas

dostępny w literaturze wynikami. Hernandez-Matamoros et al. (2020) wykazał istnienie relacji pomiędzy rozprzestrzenianiem się wirusa a innymi zmiennymi charakterystycznymi dla krajów należących do tego samego regionu geograficznego. García-Cremades et al. (2021) uwzględnił nieregularny komponent w predykcji ARIMA, którym była mobilność, co poprawiło dokładność. Dało to możliwości do włączenia innego komponentu nieregularności – szczepień. Hybrydowe rozwinięcia modelu ARIMA i SEIRD umożliwiło również poprawę predykcji, dając precyzyjną prognozę w dwumiesięcznym horyzoncie predykcji. Model ARIMA porównywano z ANN, gdzie ARIMA okazała się bardziej precyzyjna niż ANN, które zaniżało wartości prognozowane względem rzeczywistych.

Do predykcji rozprzestrzeniania się wirusa wykorzystywano technologie oparte na AI i ML. Modele takie radykalnie poprawiły dokładność predykcji, śledzenie kontaktów, prace nad lekiem i szczepionką (Lalmuanawma et al. (2020)). Wadami modeli DL są przede wszystkim niepowtarzalność otrzymanych wyników oraz konieczność stałego i czasochłonnego trenowania sieci w celu predykcji. Każda zmiana horyzontu predykcji wymaga zmiany danych wejściowych i budowanie modelu na nowo. Sytuacja ma się nieco lepiej w przypadku RNN. Współczesna dostępność do mocy obliczeniowej w technologiach chmurowych umożliwia dalsze prace nad DL i predykcją zakażeń.

W niniejszej rozprawie zaproponowano 3 metody predykcyjne szeregów czasowych zakażeń wirusem w różnych krajach. Metoda multiregresji z pseudoinwersją Moore'a – Penrose'a uwzględnia szczepienia poprzez zastosowanie zaaprobowanego kryterium gęstości szczepień. Metoda oparta na łańcuchach Markowa oparta jest na prawdopodobieństwie przejścia pomiędzy binami - danymi sklasyfikowanymi w 5 binach i tygodniowych oknach czasowych. Kolejna proponowana metoda n-najbliższych sąsiadów ze średnią krocząca opiera się na informacji o zakażeniach z n-krajów o podobnych znormalizowanych przebiegach szeregów czasowych zakażeń. Każda z wymienionych metod wymaga zastosowania rzetelnych danych wejściowych do predykcji. W zaimplementowanych skryptach realizujących algorytmy wymienionych metod, dane pobierane są online z repozytorium danych. Każda z proponowanych metod ma swoje ograniczenia.

Przeprowadzono badania dokładności metod w zależności od długości horyzontu predykcji. Badaniu poddano podobieństwo dokładności predykcji dla krajów z tego samego regionu geograficznego. Zbadano uwzględnienie szczepień na dokładność predykcji. Dokładność metod

zbadano w różnych fazach pandemii na ekstremalnym przykładzie fali zakażeń w Indiach. Sprawdzono również jak zmienia się dokładność predykcji w zależności od wyboru liczby krajów o podobnych przebiegach. Dla każdej z prezentowanych metod udokumentowano główne wykresy i dane z co najmniej jednego uruchomienia, w przypadku pozostałych uruchomień zaprezentowano najistotniejsze obserwacje. Podczas badań odnotowano ciekawe obserwacje, które zostaną przedstawione w dalszej części.

Badanie rozpoczęto od analizy dokładności predykcji w metodzie n-najbliższych sąsiadów ze średnią kroczącą w zależności od liczby sąsiadów i horyzontu predykcji. Wynik umożliwiłby znalezienie optymalnej wartości parametru n, który w pełni pozwoli wykorzystać dokładność tej metody. Najdokładniejsze prognozy osiągnięto z parametrem  $n=7$ . Zmniejszenie liczby sąsiadów skutkowało większym błędem niż ich zwiększenie. Zwiększenie horyzontu predykcji skutkowało w każdym przypadku obniżeniem wartości MAPE. Biorąc pod uwagę powyższe wnioski, do kolejnych badań metodą n-najbliższych sąsiadów ze średnią kroczącą zastosowano wartość parametru  $n=7$ .

Badanie dokładności predykcji zaproponowanych metod w zależności od horyzontu predykcji przeprowadzono na danych z jednej z największych fal zakażeń jaka miała miejsce w Indiach. Wybór był nieprzypadkowy. Ekstremalne wartości danych stawiały duże wyzwanie dla algorytmów. W końcu powinny one działać w każdych sytuacjach, co potwierdzono w innym badaniu. Spośród badanych metod, niezależnie od długości badanego horyzontu predykcji lub fazy pandemii, najbardziej dokładna okazała się multiregresja z pseudoinwersją Moore'a – Penrose'a. W najdłuższym badanym horyzoncie predykcji wynoszącym 3 tygodnie dla 1 fazy pandemii w Indiach multiregresja osiągnęła dokładność MAPE 0,08, podczas gdy metoda n-najbliższych sąsiadów z optymalnym parametrem  $n=7$  osiągnęła dokładność MAPE na poziomie 0,43. Metoda oparta na łańcuchach Markowa ze względu na działanie w tygodniowych oknach czasowych została badana w tygodniowym horyzoncie predykcji osiągając wynik dokładności MAPE 0,24, jest to wynik zbliżony do metody n-najbliższych sąsiadów (0,21). Każda z metod badanych powyżej tygodniowego horyzontu predykcji odnotowała spadek dokładności wraz ze wzrostem horyzontu predykcji. W przypadku badania dla horyzontu predykcji poniżej tygodnia (1 i 3 dni) najgorzej wypadła metoda oparta na łańcuchach Markowa (MAPE odpowiednio 0,61 i 0,32). Przyzwoite wyniki dokładności odnotowano dla metody n-najbliższych sąsiadów ze średnią kroczącą, MAPE kształtowało się odpowiednio na poziomie 0,16 i 0,18. Przydatny do analizy

graficznej prognozy tą metodą okazał się zaproponowany format wykresu, w którym wartości prognozowane stanowiły przedłużenie wartości rzeczywistych. Bezkonkurencyjna pod względem dokładności predykcji była metoda multiregresji z pseudoinwersją Moore'a – Penrose'a osiągając wartości MAPE  $\sim 0,01$  i  $\sim 0,007$ .

Wszystkie opracowane metody porównano również w badaniu wpływu położenia geograficznego ( $hp=7$ ) oraz w badaniu wpływu fazy pandemii ( $hp=7$ ) na dokładność predykcji w tygodniowym horyzoncie predykcji. Porównano uśrednione wyniki dla każdej z metod dla badań wszystkich krajów. Ponownie najdokładniejsza była multiregresja z pseudoinwersją Moore'a – Penrose'a (odpowiednio 0,007 i 0,013), następnie metoda n-najbliższych sąsiadów (odpowiednio 0,018 i 0,10) i metoda oparta na łańcuchach Markowa (0,31), która nie była porównywana w drugim badaniu.

Następnie zbadano wpływ uwzględnienia szczepień na dokładność predykcji. Jediną metodą uwzględniającą szczepienia jest multiregresja z pseudoinwersją Moore'a Penrose'a. W przypadku pozostałych metod nie uwzględniano szczepień. Metoda oparta na łańcuchach Markowa działa na jednym szeregu czasowym prognozowanego kraju. Metoda n-najbliższych sąsiadów uwzględnia znormalizowane wartości skumulowanych dobowych zakażeń. Oczywiście, istnieje możliwość włączenie czynnika szczepień również w tych metodach, co może być dokonane jako kontynuacja badań. Na potrzeby niniejszej rozprawy potrzebne było badanie dokładności metod w uproszczonej formie. Uwzględnienie szczepień w badanej metodzie odbyło się poprzez dobór krajów według zdefiniowanego kryterium gęstości szczepień. Wybrano kraj o wysokiej i kraj o niskiej gęstości szczepień przypadającej na dzień początku okna uczenia. Dla każdego z tych krajów wybrano do predykcji kraje podobne pod względem gęstości szczepień, których szeregi czasowe skumulowanych dobowych zakażeń stanowiły zmienne niezależne. Badanie wykonano w 2 różnych zakresach dat: kiedy pojawiły się dostępne szczepienia oraz kilka miesięcy później, kiedy szczepienia już były powszechnie dostępne na rynku. Wyniki badania w pierwszym oknie czasowym wykazały większą dokładność predykcji dla kraju z większą gęstością szczepień, natomiast w drugim oknie czasowym wykazały odmienne wyniki, najdokładniejsza predykcja była dla kraju z najmniejszą gęstością szczepień. Świadczy to o tym, że uwzględnienie szczepień w badanej metodzie poprzez przyjęte kryterium gęstości szczepień nie ma wpływu na dokładność predykcji. Kolejną ważną obserwację w kontekście szczepień zaobserwowano badając wpływ dokładności predykcji na położenie geograficzne krajów. Badane kraje afrykańskie miały

średnią gęstość szczepień wynoszącą 269, podczas gdy średnie gęstości szczepień dla krajów z Europy, Azji i Ameryki Południowej mieściły się w zakresie 1311-1400. Średnie MAPE uwzględnionych metod dla krajów afrykańskich – najslabiej zaszczepionych według przyjętego kryterium – wyniosło 0,4, podczas gdy kraje zgrupowane w pozostałych kontynentach miały ponad dwukrotnie wyższą dokładność mieszczącą się w zakresie MAPE 0,14-0,17. Biorąc pod uwagę obserwację z badania regionów geograficznych nie można jednoznacznie stwierdzić, że dokładność predykcji ma związek z gęstością szczepień, ponieważ w przypadku krajów afrykańskich stwierdzono problem z jakością raportowanych danych. Mógł być to czynnik istotnie wpływający na dokładność predykcji.

Kontynuując spostrzeżenia z badania wpływu położenia geograficznego na dokładność predykcji, Azja miała najdokładniejszą prognozę dla metody opartej na łańcuchach Markowa (MAPE = 0,34) i multiregresji z pseudoinwersją Moore'a – Penrose'a (MAPE = 0,002), w przypadku metody n-najbliższych sąsiadów ze średnią krocząca najdokładniejszą predykcję odnotowała Europa (MAPE = 0,03). Kraje afrykańskie miały najmniej dokładną prognozę we wszystkich badanych metodach. Rekordzistą najmniej precyzyjnej prognozy okazało się Malawi, które w metodzie opartej na łańcuchach Markowa osiągnęło MAPE na poziomie 0,98. Można przyznać, że położenie geograficzne, a co za tym idzie zróżnicowanie kulturowe, społeczne i gospodarcze, ma istotny wpływ na dokładność prognozy skumulowanych dobowych zakażeń wykonanej za pomocą wspomnianych metod.

Dwie z zaproponowanych metod były badane pod względem dokładności prognoz rozprzestrzeniania się wirusa w różnych fazach pandemii. Metoda oparta na łańcuchach Markowa została pominięta w tym badaniu, ponieważ jej dokładność w tygodniowym horyzoncie predykcji była znacznie niższa niż pozostałych oraz przede wszystkim działa ona w tygodniowym horyzoncie predykcji. Zdefiniowano 4 fazy w oparciu o wspomnianą już falę zakażeń w Indiach. Początek fali zakażeń okazał się najtrudniejszym dz dwutygodniowych horyzontów predykcji, stwierdzono najniższą dokładność predykcji (MAPE = 0,30) dla metody n-najbliższych sąsiadów ze średnią krocząca, dla multiregresji z pseudoinwersją Moore'a – Penrose'a najmniejsza dokładność dla tej fazy wyniosła MAPE = 0,04. Bardzo dobrze poradziły sobie obie metody z przejściem przez pik zakażeń. Była to najważniejsza z obserwacji. Nagły zwrot w postaci pierwszych niższych dobowych przyrostów zakażeń zostały uwzględnione w oknach uczenia. Obserwacja ta świadczy o dokładności zastosowania zaproponowanych metod do predykcji

rozprzestrzeniania się wirusa krótko po wprowadzeniu zmian w polityce obostrzeń, a tym samym badania dokładności podejmowanych środków w zapobieganiu rozprzestrzeniania się wirusa. Dokładność predykcji wzrastała w kolejnych badanych fazach: końca fali zakażeń oraz stabilizacji dobowych przyrostów zakażeń z tendencją wzrostową pod koniec okna uczenia. Reasumując, o ile predykcję na poziomie MAPE 0,3 można nazwać dokładną, o tyle obie metody są dokładne niezależnie od fazy pandemii w danym kraju.

W trakcie badań odnotowano szereg interesujących spostrzeżeń związanych z proponowanymi metodami i dostępnymi danymi. Stosując przedstawione metody predykcji można było przewidzieć wzrost zakażeń w Indiach pomiędzy kwietniem a czerwcem 2021 r. Szczególnie dokładna w tym konkretnym przypadku okazała się metoda oparta na multiregresji i pseudoinwersji Moore'a – Penrose'a. Rzetelne informacje o spodziewanym wzroście zakażeń umożliwiłyby przygotowanie się służby zdrowia, zabezpieczenie zasobów, łóżek szpitalnych, zapasu tlenu, zagospodarowania ciał zmarłych, zrobienie zapasów leków, opracowanie komunikacji społecznej.

Metoda oparta na łańcuchach Markowa nie wymaga danych innych krajów niż badanego. Bazuje bowiem ona na prawdopodobieństwie przejść między danymi sklasyfikowanymi w binach. Metoda ta na pewno będzie rozwijana w innej dziedzinie, np. finansach lub telematyce. Predykcja z wykorzystaniem tej metody jest dokładna w bardzo krótkim okresie czasu, zalecany  $hp=1$ . Wyzwaniem jest rekurencyjne wywołanie metody w  $hp=1$ , gdzie wartość prognozowana będzie wchodziła do okna uczenia w kolejnym wywołaniu metody. Umożliwi to predykcję w nieograniczonym horyzoncie predykcji.

Rozwinięciem metody multiregresji z pseudoinwersją Moore'a-Penrose'a mogłoby być wyłączenie badania gęstości szczepień. Umożliwiłoby to badanie predykcji w zakresach dat, zwłaszcza pierwszych miesiącach pandemii, kiedy nie było jeszcze dostępnych szczepionek.

Z kolei rozwinięciem pozostałych dwóch metod byłoby uwzględnienie szczepień, np. w postaci zmiennych współczynników (łańcuchy Markowa) lub jako dodatkowe kryterium doboru n-sąsiadów.

Szybkość wykonywania obliczeń w predykcji jest atutem rozważanych metod – prawie wszystkie symulacje wykonywane są w czasie poniżej 5 sekund, najdłużej wykonywana jest predykcja oparta na podobieństwie n-sąsiadów i związana z tym analiza korelacji wzajemnej krajów. Obliczenia dla parametru n-sąsiadów  $n>9$  i horyzontu predykcji  $hp=14$  zajmuje około 4

minut. Szybkość wykonywanych prognoz oraz niskie zapotrzebowanie na moc obliczeniową umożliwia wdrożenie metod na sprzętowych rozwiązaniach, np. telemetrycznych, pojazdach autonomicznych oraz innych urządzeniach Internetu rzeczy.

Zaletą jest również pełna powtarzalność wyników. Wszystkie wyniki opierające się na tym samym zestawie danych wejściowych i wywołane z tymi samymi parametrami zawsze zwracają te same wyniki. W przypadku analizowanych rozwiązań opartych na Deep Learning nie uzyskiwano powtarzalnych wyników.

Dostęp do danych, testowanie na obecność wirusa i sposób raportowania wirusa jest niespójny w różnych krajach, co wpływa na jakość predykcji. Repozytoria danych nie zawierają informacji, kiedy i jaki kraj zmienił podejście w tym temacie. Ustalenie zmian dynamiki szeregów czasowych związanych z COVID-19 możliwe jest za pomocą modeli opartych na równaniach różniczkowych, jednak nieznane są źródła zmian ich dynamiki. Takimi źródłami mogą być również zmiany w polityce obostrzeń lub zmiany meteorologiczne. Instytucje monitorujące stan rozwoju epidemii powinny zatem raportować i testować obecność wirusa w sposób spójny na poziomie kraju.

W trakcie badań potwierdzono związek zamożności badanych krajów ze szczepieniami populacji. Kraje z Afryki, powszechnie uznawane za słabo rozwinięte gospodarczo, miały radykalnie mniejszą gęstość szczepień. Gęstość szczepień, zwłaszcza w początkowym okresie dostępności do szczepionki, była wysoka dla krajów uznawanych za rozwinięte gospodarczo, w późniejszym czasie wartość ta była wysoka również dla krajów zamożnych i o małej powierzchni (np. San Marino). Gęstość szczepień utrzymywała niską pozycję w większości krajów z Afryki.

Analiza wyników prognozowania inflacji za pomocą zaproponowanych metod regresji wielomianowej i metody wygładzającej splajnami wykazała znaczące różnice w dokładności tych podejść. Regresja wielomianowa, choć dokładna w modelowaniu zależności nieliniowych, wykazała ograniczenia w kontekście adaptacji do zmieniających się warunków rynkowych. Z kolei metoda splajnów, dzięki swojej elastyczności w dopasowywaniu lokalnych zmienności danych, wykazała większą efektywność w przewidywaniu krótkoterminowych fluktuacji inflacyjnych. Te wyniki podkreślają, że choć oba podejścia mają swoje zalety, metoda splajnów może być bardziej odpowiednia dla dynamicznie zmieniającego się środowiska gospodarczego.

Optymalizacja parametru wygładzającego splajnów  $p$  przez iterację wypadła słabiej niż domyślna metoda funkcji dopasowania w MATLAB. To może sugerować, że automatyczne

algorytmy MATLAB są wysoce dokładne w znalezieniu optymalnego balansu między dopasowaniem a gładkością modelu, nawet bez dodatkowej ręcznej optymalizacji. Może to również wskazywać na konieczność dalszych badań w celu zrozumienia, dlaczego metody ręcznej optymalizacji nie przyniosły oczekiwanych korzyści w poprawie prognoz.

Zastosowanie korekty ECM do modelu regresji wielomianowej przyniosło niewielką poprawę dokładności prognoz, co wyraziło się w zmniejszeniu błędu MSE z 1,71 do 1,62 oraz RMSE z 0,98 do 0,94, przy równoczesnym niewielkim wzroście wartości współczynnika  $R^2$  z 0,91 do 0,92. To wskazuje na to, że korekta ECM może mieć pozytywny, choć ograniczony wpływ na zdolność modelu do prognozowania inflacji, zwłaszcza w kontekście danych wykazujących relacje równowagi długoterminowej. Jednakże, zastosowanie tej samej korekty do modelu regresji wygładzającej splajnem z optymalizacją nie przyniosło oczekiwanych korzyści, gdyż średnie wartości MSE i RMSE pozostały niemal niezmiennie (MSE z 1,71 do 1,69 i RMSE z 1,02 do 1,01), z zachowaniem wartości  $R^2$  na poziomie 0,88. To może sugerować, że w przypadku danych o mniejszej strukturze równowagi długoterminowej, korzyści wynikające z korekty ECM nie są wyraźne. Zaskakująco, najlepsze wyniki osiągnięto dla modelu regresji wygładzającej splajnem bez optymalizacji, gdzie osiągnięto najniższe wartości MSE i RMSE (odpowiednio 1,47 i 0,93), co sugeruje wysoką dokładność tego podejścia w prognozowaniu zjawisk inflacyjnych. Wartości  $R^2$  dla tego modelu wyniosły 0,90, co wskazuje na silne dopasowanie modelu do danych.

Te wyniki podkreślają znaczenie wyboru odpowiedniej metody modelowania w kontekście specyfiki badanych danych i mogą wskazywać na ograniczenia zastosowania korekty ECM w odniesieniu do niektórych typów modeli regresyjnych, zwłaszcza gdy dane nie wykazują wyraźnych relacji równowagi długoterminowej.

Istotny okazał się wybór narzędzi programistycznych do implementacji algorytmów i akwizycji danych. Wybór był pomiędzy komercyjnym narzędziem MATLAB a językiem Python i darmowymi dostępnymi bibliotekami. W przypadku algorytmów „pandemicznych” wybór padł na język Python i stworzone w nim biblioteki. Dostępność narzędzi umożliwi dalsze kontynuowanie prac badawczych bez konieczności zakupu kosztownej licencji komercyjnej MATLAB lub po wygaśnięciu licencji edukacyjnej. Kolejnym argumentem jest mnogość dostępnych bibliotek, w tym dedykowanej dla modelowania pandemii - covsirphy („covsirphy · PyPI”, <https://pypi.org/project/covsirphy/> (2022)). W przypadku predykcji inflacji wykorzystano środowisko MATLAB wraz z jego rozszerzeniami dostępnymi w ramach licencji studenckiej.

Programowanie w środowisku MATLAB umożliwiło wykorzystanie jego potencjału. Wbudowane funkcje optymalizacyjne w regresji wygładzania splajnem okazały się dokładniejsze niż ręczna optymalizacja parametrem wygładzania  $p$ . Pewnym mankamentem może okazać się kontynuacja badań w tym środowisku po wygaśnięciu licencji i utracie dostępu do rozszerzeń.

W ramach przeprowadzonych badań udało się osiągnąć rozwiązać problem badawczy dotyczący prognozowania inflacji. Model regresji wygładzającej splajnem bez optymalizacji zapewnił najniższe wartości błędów, takie jak MSE i RMSE, co wskazuje na jego wysoką dokładność i potencjał do dalszej rozbudowy. Metody bazują na jednym wskaźniku CPI, wykazały się elastycznością i możliwością zastosowania dla dowolnego kraju, co jest istotne w kontekście globalnej zmienności warunków ekonomicznych. Wykazano również, że metody badawcze w sposób dostateczny reagowały na szoki zewnętrzne, wykazując odporność na takie czynniki jak kryzysy finansowe czy zmiany polityczne. Co istotne, wyniki prognoz okazały się zadowalające, dostarczając solidnej podstawy dla podejmowania decyzji ekonomicznych, co potwierdza ich praktyczną przydatność i znaczenie.

Podsumowując, w niniejszej rozprawie potwierdzono stawianą hipotezę badawczą. Zaproponowano 3 odmienne metody predykcji rozprzestrzeniania się wirusa na poziomie kraju oraz 2 metody prognozowania inflacji rozwinięte o korekcję i optymalizację. Metody prognozy szeregów „pandemicznych” dają precyzyjne prognozy w 3 tygodniowych horyzontach predykcji, niezależnie od fazy fali zakażeń i położenia geograficznego badanego kraju. Metody prognozy szeregów inflacyjnych dają dobre prognozy w jednomiesięcznych horyzontach predykcji. Wykazano również, że metody z powodzeniem mogą być wykorzystywane do badania wpływu wprowadzanych polityk zapobiegania rozprzestrzeniania się wirusa.

Autor uważa, że z pokorą wobec nauki wniósł swój skromny wkład w dyscyplinę informatyki technicznej i telekomunikacji.

1. Interdyscyplinarne podejście do problemów społecznych i zdrowia publicznego oraz problemów ekonomicznych w zastosowaniu inteligencji obliczeniowej w predykcji w szeregach czasowych procesów złożonych. Dokładna predykcja umożliwi podjęcie decyzji w nowych obszarach: epidemiologicznym, ekonomicznym lub zarządzania kryzysowego.

2. Zaproponowano autorskie metody predykcji w szeregach czasowych mogą mieć szerokie zastosowanie wykraczające poza domenę pandemii i inflacji. Mogą z powodzeniem być wdrożone we wskazywanych przez autora licznych dziedzinach.
3. Opracowanie metody korekcji błędu ECM umożliwiającą zwiększenie dokładności prognozowanych wartości. Metoda wykorzystuje podejście prognozowania przyszłej wartości polegające na uczeniu się na błędach prognoz wartości historycznych. Innymi słowy, oprócz prognozowanej wartości szeregu czasowego prognozuje się również błąd. Metoda ma istotny wkład w predykcję w szeregach czasowych i podkreśla wykorzystanie metod sztucznej inteligencji do osiągnięcia dokładniejszych wyników.
4. Autor dokonał analizy porównawczej modeli predykcyjnych w szeregach czasowych. Szczegółowe porównanie umożliwia zrozumienie, które modele i metody sztucznej inteligencji są bardziej dokładne w określonych scenariuszach. Wskazywane metody rozwoju i optymalizacji umożliwiają poszerzenie narzędzi dostępnych w analizie szeregów czasowych.
5. Predykcja w szeregach czasowych inflacji z mierzalnym błędem opartym na wartościach historycznych daje możliwość podjęcia decyzji inwestycyjnych w mikro i makroekonomicznej skali oraz umożliwia przeprowadzenie rzetelnych analiz ekonomicznych.
6. W pracy podkreślono znaczenie i jakość danych oraz istotę badań związanych z ich gromadzeniem i przetwarzaniem, co ma znaczący wpływ na dokładność predykcji.

## 5.2 Dalsze kierunki badań

W trakcie pracy nad niniejszą rozprawą doktorską zaproponowano dalsze kierunki badań. Jednym z nich jest doskonalenie algorytmów poprzez opracowanie współczynników optymalizujących, np. redukujących załamanie na styku wykresów szeregów czasowych rzeczywistego i prognozowanego. Innym kierunkiem jest uwzględnienie szczepień w metodzie n-najbliższych sąsiadów.

Ciekawym kierunkiem badań jest prognozowanie wymienionymi metodami innych szeregów czasowych związanych z wirusem SARS-CoV-2, np. zgonów, ozdowień lub powtórnych zakażeń.

Autor niniejszej rozprawy ma również w planach badanie proponowanymi metodami (po ich adaptacji) szeregów czasowych z innych dziedzin, np. finansów, teledzienników lub militarnych.

Badanie zmodyfikowanej metody opartej na łańcuchach Markowa do jej rekurencyjnego wywołania umożliwiłoby wykonanie predykcji przekraczającej jednodniowy horyzont predykcji.

Należałoby zbadać możliwość wdrożenia proponowanych algorytmów w reprogramowalnych układach cyfrowych stosowanych np. w urządzeniach Internetu rzeczy do predykcji szeregów czasowych danych z sensorów (np. pojazdy autonomiczne).

Innym ciekawym kierunkiem jest zaimplementowanie mikro usług sieciowych wykorzystujących proponowane metody predykcji. Umożliwiłoby to predykcję rozprzestrzeniania się wirusa przez szeroką grupę badaczy.

W trakcie pracy nad niniejszą rozprawą doktorską, skupiającą się na modelowaniu i prognozowaniu inflacji, wyłoniło się kilka istotnych obszarów dla dalszych badań. Kluczowym kierunkiem jest doskonalenie istniejących algorytmów, ze szczególnym uwzględnieniem opracowania współczynników optymalizujących. Takie współczynniki mogłyby pomóc w zmniejszaniu różnic między rzeczywistymi a prognozowanymi szeregami czasowymi inflacji, zwiększając tym samym dokładność i wiarygodność prognoz. Idąc za literaturą, istotne wydaje się badanie metod opartych na wielomianach ułamkowych. Takie podejście może oferować nowe możliwości w modelowaniu zjawisk inflacyjnych, szczególnie w kontekście złożonych i nieregularnych wzorców danych. Kolejnym ważnym kierunkiem jest eksploracja zastosowania metod hybrydowych, łączących różne techniki modelowania.

Ponadto, warto badać wpływ niepewności gospodarczych i kryzysów, takich jak pandemia COVID-19, na modele prognozowania inflacji. Taka analiza pozwoliłaby ocenić, jak modele reagują na gwałtowne zmiany w gospodarce i w jaki sposób można je dostosować, aby były bardziej odporne na takie szoki.

Najważniejszym kierunkiem dalszych badań jest rozważanie zastosowania opracowanych metod prognozowania do innych kluczowych wskaźników „pandemicznych” i ekonomicznych. Holistyczne podejście do analizy i prognozowania istotnych społecznie szeregów czasowych może dostarczyć bardziej kompleksowego obrazu i lepiej informować decyzje polityczne i gospodarcze.

# Bibliografia

1. Adiguzel, F. S., Cansunar, A., & Corekcioglu, G. (2020). Truth or dare? Detecting systematic manipulation of covid-19 statistics. *Journal of Political Institutions and Political Economy*, 1(4), 543-557.
2. Ala'raj, M., Majdalawieh, M., & Nizamuddin, N. (2021). Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections. *Infectious Disease Modelling*, 6, 98-111.
3. Alenezi, M. N., Al-Anzi, F. S., Alabdulrazzaq, H., Alhusaini, A., & Al-Anzi, A. F. (2021). A study on the efficiency of the estimation models of COVID-19. *Results in Physics*, 26, 104370.
4. Atkeson, A., & Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review*, 25(1), 2-11.
5. Bańbura, M., & Bobeica, E. (2023). Does the Phillips curve help to forecast euro area inflation? *International Journal of Forecasting*, 39(1), 364-390.
6. Bar-Yam, Y. (2019). *Dynamics of complex systems*. CRC Press.
7. Barabási, A. L. (2016). *Network Science*. Cambridge University Press, Cambridge.
8. Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., & Koenigstein, N. (2023). Forecasting CPI inflation components with hierarchical recurrent neural networks. *International Journal of Forecasting*, 39(3), 1145-1162.
9. Bashir, M. F., Ma, B., Bilal, Komal, B., Bashir, M. A., Tan, D., & Bashir, M. (2020). Correlation between climate indicators and COVID-19 pandemic in New York, USA. *The Science of the total environment*, 728, 138835. <https://doi.org/10.1016/j.scitotenv.2020.138835>
10. Bashir, M. F., Ma, B., & Shahzad, L. (2020). A brief review of socio-economic and environmental impact of Covid-19. *Air Quality, Atmosphere & Health*, 13(12), 1403-1409.
11. Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markow chains. *The annals of mathematical statistics*, 37(6), 1554-1563.
12. Ben-Israel, A. (2002). The moore of the moore-penrose inverse. *The Electronic Journal of Linear Algebra*, 9, 150-157.

13. Binder, H., Sauerbrei, W., & Royston, P. (2013). Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. *Statistics in Medicine*, 32(13), 2262-2277
14. Bhandari, P. (2019). Analysis of prediction models in spread of ebola virus disease (Doctoral dissertation, Thesis, Deakin University).
15. Bhatnagar, M. R. (2020). COVID-19: Mathematical modeling and predictions. ResearchGate, DOI: 10.13140/RG.2.2.29541.96488
16. Bobkowska, J. (2014). Wady i zalety metody prognozowania zachowania złożonych procesów opartej na metodzie GMDH z funkcjami wrażliwości. *Pomiary Automatyka Kontrola*, 60(12), 1136-1139
17. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
18. Breitung, J., & Roling, C. (2015). Forecasting inflation rates using daily data: A nonparametric MIDAS approach. *Journal of Forecasting*, 34(7), 588-603.
19. Cano-Pérez, E., Torres-Pacheco, J., Fragozo-Ramos, M. C., García-Díaz, G., Montalvo-Varela, E., & Pozo-Palacios, J. C. (2020). Negative Correlation between Altitude and COVID-19 Pandemic in Colombia: A Preliminary Report. *The American journal of tropical medicine and hygiene*, 103(6), 2347–2349. <https://doi.org/10.4269/ajtmh.20-1027>
20. Caraka, R. E., & Sugiyarto, W. (2016). Inflation Rate Modelling in Indonesia. *None*, 15(2), 194861.
21. Chang, H. J. (2017). Estimation of basic reproduction number of the Middle East respiratory syndrome coronavirus (MERS-CoV) during the outbreak in South Korea, 2015. *Biomedical engineering online*, 16(1), 1-11.
22. Chaudhry, I. S., Iqbal, A., Umar, M., & Faheem, M. (2021). Impact of Monetary Policy on Inflation and Investment in Pakistan: A Time Series Analysis. *Journal of Accounting and Finance in Emerging Economies*, 7(4), 931-943.
23. Cheval, S., Mihai Adamescu, C., Georgiadis, T., Herrnegger, M., Piticar, A., & Legates, D. R. (2020). Observed and potential impacts of the COVID-19 pandemic on the environment. *International journal of environmental research and public health*, 17(11), 4140.

24. Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 135, 109864.
25. Ciżkowicz, P., & Rzońca, A. (2011). Koszty inflacji–przegląd piśmiennictwa. *Ekonomista*, 395.
26. Coibion, O., Gorodnichenko, Y., Kumar, S., & Pedemonte, M. (2020). Inflation expectations as a policy tool?. *Journal of International Economics*, 124, 103297.
27. Cooper, I., Mondal, A., & Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139, 110057.
28. D'Acunto, F., Hoang, D., Paloviita, M., & Weber, M. (2021). Human frictions in the transmission of economic policies (No. w29279). National Bureau of Economic Research.
29. de Jong, M. C. (1995). Mathematical modelling in veterinary epidemiology: why model building is important. *Preventive veterinary medicine*, 25(2), 183-193.
30. de León, U. A. P., Avila-Vales, E., & Huang, K. L. (2022). Modeling COVID-19 dynamic using a two-strain model with vaccination. *Chaos, Solitons & Fractals*, 157, 111927.
31. de León, U. A. P., Pérez, Á. G., & Avila-Vales, E. (2020). An SEIARD epidemic model for COVID-19 in Mexico: mathematical analysis and state-level forecast. *Chaos, Solitons & Fractals*, 140, 110165.
32. Drozdowicz-Bieć, M. (2011). Reasons why Poland avoided the 2007-2009 Recession. *Prace i Materiały Instytutu Rozwoju Gospodarczego SGH*, 86, 39-66.
33. Dharma, F., Shabrina, S., Noviana, A., Tahir, M., Hendrastuty, N., & Wahyono, W. (2020). Prediction of Indonesian inflation rate using regression model based on genetic algorithms. *Jurnal Online Informatika*, 5(1), 45-52.
34. Dutta, S., & Bandyopadhyay, S. K. (2020). Machine learning approach for confirmation of covid-19 cases: Positive, negative, death and release. MedRxiv.
35. Eifan, S. A., Nour, I., Hanif, A., Zamzam, A. M., & AlJohani, S. M. (2017). A pandemic risk assessment of Middle East respiratory syndrome coronavirus (MERS-CoV) in Saudi Arabia. *Saudi journal of biological sciences*, 24(7), 1631-1638.
36. Er, P. H., Tugcu, C. T., & Coban, O. (2014). Investigating the link between savings, inflation and economic growth: An ARDL analysis for the case of Turkey. *Journal of Economics Finance and Accounting*, 1(2), 81-90.

37. Fibriyani, V., & Chamidah, N. (2021, February). Prediction of inflation rate in Indonesia using local polynomial estimator for time series data. In *Journal of Physics: Conference Series* (Vol. 1776, No. 1, p. 012065). IOP Publishing.
38. Fonseca, D., García-Peñalvo, F. J., & Camba, J. D. (2021). New methods and technologies for enhancing usability and accessibility of educational data. *Universal access in the information society*, 20(3), 421-427.
39. García-Cremades, S., Morales-García, J., Hernández-Sanjaime, R., Martínez-España, R., Bueno-Crespo, A., Hernández-Orallo, E., ... & Cecilia, J. M. (2021). Enhancing the prediction of COVID-19 evolution by combining models and data sources.
40. Gervasi, V., Marcon, A., Bellini, S., & Guberti, V. (2019). Evaluation of the efficiency of active and passive surveillance in the detection of African swine fever in wild boar. *Veterinary Sciences*, 7(1), 5.
41. Gibney, E. (2020). Coronavirus lockdowns have changed the way Earth moves. *Nature*, 580(7802), 176-178.
42. Gleick, J. (2008). *Chaos: Making a new science*. Penguin.
43. Grabia, T. (2022). The impact of the coronavirus pandemic crisis on inflation and interest rate policy in Poland. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 66(1), 47-58.
44. Hadley, E., Rhea, S., Jones, K., Li, L., Stoner, M., & Bobashev, G. (2022). Enhancing the prediction of hospitalization from a COVID-19 agent-based model: A Bayesian method for model parameter estimation. *Plos one*, 17(3), e0264704.
45. Han, E., Tan, M. M. J., Turk, E., Sridhar, D., Leung, G. M., Shibuya, K., ... & Legido-Quigley, H. (2020). Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. *The Lancet*, 396(10261), 1525-1534.
46. Hai, T. N., Van, Q. N., & Thi Tuyet, M. N. (2021). Digital transformation: Opportunities and challenges for leaders in the emerging countries in response to COVID-19 pandemic. *Emerging Science Journal*, 5, 21-36.
47. Harris, P., & Moss, D. (2020). Covid, pandemics, plague and public affairs: Lessons from history. *Journal of public affairs*, 20(4), e2548. <https://doi.org/10.1002/pa.2548>

48. Hassani, H., Soofi, A. S., & Zhigljavsky, A. (2013). Predicting inflation dynamics with singular spectrum analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 176(3), 743-760.
49. Hernandez-Matamoros, A., Fujita, H., Hayashi, T., & Perez-Meana, H. (2020). Forecasting of COVID19 per regions using ARIMA models and polynomial functions. *Applied soft computing*, 96, 106610.
50. Higazy, M. (2020). Novel fractional order SIDARTHE mathematical model of COVID-19 pandemic. *Chaos, Solitons & Fractals*, 138, 110007.
51. Huang, G., Pan, Q., Zhao, S., Gao, Y., & Gao, X. (2020). Prediction of COVID-19 outbreak in China and optimal return date for university students based on propagation dynamics. *Journal of Shanghai Jiaotong University (Science)*, 25(2), 140-146.
52. Hoang, V. T., Al-Tawfiq, J. A., & Gautret, P. (2020). The Tokyo Olympic Games and the risk of COVID-19. *Current tropical medicine reports*, 7(4), 126-132.
53. Huremović, D. (2019). Brief history of pandemics (pandemics throughout history). *In Psychiatry of pandemics* (pp. 7-35). Springer, Cham
54. Iivari, N., Sharma, S., & Ventä-Olkkonen, L. (2020). Digital transformation of everyday life—How COVID-19 pandemic transformed the basic education of the young generation and why information management research should care?. *International Journal of Information Management*, 55, 102183.
55. Ivakhnenko A.G., Zaichenko J.P., Dimitrov V.D. Priniatije resenij na osnove samoorganizacii. Sovetskoe Radio, Moskva 1976.
56. Jimi, H., & Hashimoto, G. (2020). Challenges of COVID-19 outbreak on the cruise ship Diamond Princess docked at Yokohama, Japan: a real-world story. *Global Health & Medicine*.
57. Katsikis, V., & Pappas, D. (2008). Fast computing of the Moore-Penrose inverse matrix. *The Electronic Journal of Linear Algebra*, 17, 637-650.
58. Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772), 700-721.
59. Kędzierski, M., Lisiakiewicz, R., Matyja, R., Możdżeń, M., Oramus, M., & Żabiński, M. (2020). Polityka zarządzania restrykcjami wprowadzonymi ze względu na epidemię

COVID-19–wielowymiarowy model analityczno--decyzyjny. Kraków: Centrum Polityk Publicznych.

60. Krueger, T. (2020). Epidemia covid-19: Analiza sytuacji bieżącej i prognozy krótkoterminowe.
61. Kudryashov, N. A., Chmykhov, M. A., & Vigdorowitsch, M. (2021). Analytical features of the SIR model and their applications to COVID-19. *Applied Mathematical Modelling*, 90, 466-473.
62. Kufel, T. (2020). ARIMA-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 15(2), 181-204.
63. Kumar, G., & Kumar, R. R. (2020). A correlation study between meteorological parameters and COVID-19 pandemic in Mumbai, India. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(6), 1735-1742.
64. Kumari, P., Singh, H. P., & Singh, S. (2021). SEIAQRDT model for the spread of novel coronavirus (COVID-19): A case study in India. *Applied Intelligence*, 51(5), 2818-2837.
65. Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system?. *European Journal for Philosophy of Science*, 3, 33-67.
66. Lalmuanawma, S., Hussain, J., & Chhakchhuak, L. (2020). Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals*, 139, 110059.
67. LePan, N. (2020). Visualizing the history of pandemics. *Visual Capitalist*, 14.
68. Liu, Z., Ciais, P., Deng, Z., Lei, R., Davis, S. J., Feng, S., ... & Schellnhuber, H. J. (2020). Near-real-time monitoring of global CO<sub>2</sub> emissions reveals the effects of the COVID-19 pandemic. *Nature communications*, 11(1), 1-12.
69. Maldeni, R., & Mascrenge, M. A. (2021, September). A Machine Learning Approach to CCPI-Based Inflation Prediction. In *Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021*, London, Volume 2 (pp. 567-575). Singapore: Springer Singapore.
70. Malki, Z., Atlam, E. S., Hassanien, A. E., Dagneu, G., Elhosseini, M. A., & Gad, I. (2020). Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. *Chaos, Solitons & Fractals*, 138, 110137.

71. Miller, J. H., & Page, S. E. (2008). Complex adaptive systems: an introduction to computational models of social life. (No Title).
72. Misztal, I. (2006). Properties of random regression models using linear splines. *Journal of Animal Breeding and Genetics*, 123(2), 74-80.
73. Moein, S., Nickaeen, N., Roointan, A., Borhani, N., Heidary, Z., Javanmard, S. H., ... & Gheisari, Y. (2021). Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan. *Scientific Reports*, 11(1), 1-9.
74. Moftakhar, L., Mozghan, S. E. I. F., & Safe, M. S. (2020). Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models. *Iranian Journal of Public Health*, 49(Suppl 1), 92.
75. Ndaïrou, F., Area, I., Nieto, J. J., & Torres, D. F. (2020). Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solitons & Fractals*, 135, 109846.
76. Nemes, L., & Kiss, A. (2021). Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication*, 5(1), 1-15.
77. Newman, M. E. J. (2010). 2010Networks: an introduction.
78. Novakovic, A., & Marshall, A. H. (2022). The CP-ABM Approach for Modelling COVID-19 Infection Dynamics and Quantifying the Effects of Non-Pharmaceutical Interventions. *Pattern Recognition*, 108790.
79. Nowaczewska, M. (2023). Sposoby ochrony oszczędności gospodarstw domowych przed inflacją na przykładzie Polski. *Zeszyty Studenckie Wydziału Ekonomicznego Uniwersytetu Gdańskiego "NASZE STUDIA"*, (13).
80. Paleologos, E. K., O'Kelly, B. C., Tang, C. S., Cornell, K., Rodríguez-Chueca, J., Abuel-Naga, H., ... & Singh, D. N. (2020). Post Covid-19 water and waste water management to protect public health and geoenvironment. *Environmental Geotechnics*, 8(3), 193-207.
81. Pathan, R. K., Biswas, M., & Khandaker, M. U. (2020). Time series prediction of COVID-19 by mutation rate analysis using recurrent neural network-based LSTM model. *Chaos, Solitons & Fractals*, 138, 110018.
82. Peng, L., Yang, W., Zhang, D., Zhuge, C., & Hong, L. (2020). Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv preprint arXiv:2002.06563.

83. Piret, J., & Boivin, G. (2021). Pandemics throughout history. *Frontiers in microbiology*, 3594.
84. Prahutama, A., & Santoso, R. (2018, May). Modelling inflation in transportation, communication and financial services using B-Spline time series model. In *Journal of Physics: Conference Series* (Vol. 1025, No. 1, p. 012104). IOP Publishing.
85. Priyono, A., Moin, A., & Putri, V. N. A. O. (2020). Identifying digital transformation paths in the business model of SMEs during the COVID-19 pandemic. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 104.
86. Próchnicki, L. (2011). Polityka monetarna Polski w świetle bezpośredniego celu inflacyjnego. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, 22, 91-115.
87. Rossman, H., Shilo, S., Meir, T., Gorfine, M., Shalit, U., & Segal, E. (2021). COVID-19 dynamics after a national immunization program in Israel. *Nature medicine*, 27(6), 1055-1061.
88. Routt, D. (2008). The economic impact of the Black Death. *EH. net Encyclopedia*.
89. Rume, T., & Islam, S. D. U. (2020). Environmental effects of COVID-19 pandemic and potential strategies of sustainability. *Heliyon*, 6(9)
90. Sadkowska-Todys, M. (2006). Wścieklizna-aktualne problemy epidemiologiczne. *Polski Przegląd Neurologiczny*, 2(1), 37-42.
91. Samui, P., Mondal, J., & Khajanchi, S. (2020). A mathematical model for COVID-19 transmission dynamics with a case study of India. *Chaos, Solitons & Fractals*, 140, 110173.
92. Sayama, H. (2015). Introduction to the modeling and analysis of complex systems. Open SUNY Textbooks.
93. Sehdev, P. S. (2002). The origin of quarantine. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 35(9), 1071-1072.
94. Sarbiński, J. (2016). Wzrost dochodów emerytów i rencistów na tle wzrostu wynagrodzeń w Polsce. *Ubezpieczenia Społeczne. Teoria i praktyka*, (1), 113-117.
95. Sharma, R. R., Kumar, M., Maheshwari, S., & Ray, K. P. (2020). EVDHM-ARIMA-based time series forecasting model and its application for COVID-19 cases. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-10.

96. Shastri, S., Singh, K., Kumar, S., Kour, P., & Mansotra, V. (2020). Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. *Chaos, Solitons & Fractals*, 140, 110227.
97. Shinkarenko, V., Hostryk, A., Shynkarenko, L., & Dolinskyi, L. (2021). A forecasting the consumer price index using time series model. In SHS Web of Conferences (Vol. 107, p. 10002). EDP Sciences.
98. Silva, P. C., Batista, P. V., Lima, H. S., Alves, M. A., Guimarães, F. G., & Silva, R. C. (2020). COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solitons & Fractals*, 139, 110088.
99. Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American philosophical society*, 106(6), 467-482.
100. Sims, J. (2020). Will the world be quieter after the pandemic. BBC Future.
101. Singh, R. K., Rani, M., Bhagavathula, A. S., Sah, R., Rodriguez-Morales, A. J., Kalita, H., ... & Kumar, P. (2020). Prediction of the COVID-19 pandemic for the top 15 affected countries: advanced autoregressive integrated moving average (ARIMA) model. *JMIR public health and surveillance*, 6(2), e19115.
102. Smith, G. D. (2002). Commentary: Behind the Broad Street pump: aetiology, epidemiology and prevention of cholera in mid-19th century Britain. *International Journal of Epidemiology*, 31(5), 920-932.
103. Somani, M., Srivastava, A. N., Gummadivalli, S. K., & Sharma, A. (2020). Indirect implications of COVID-19 towards sustainable environment: an investigation in Indian context. *Bioresource Technology Reports*, 11, 100491.
104. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). In: World Health Organization/Newsroom [website]. Geneva: World Health Organization; 2020 ([https://www.who.int/newsroom/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/newsroom/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov))) Stan witryny na dzień 29 luty 2020 r..

105. Steuer, R., Kurths, J., Daub, C. O., Weise, J., & Selbig, J. (2002). The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics*, 18(suppl\_2), S231-S240.
106. Stock, J. H., & Watson, M. W. (2008). Phillips curve inflation forecasts.
107. Strogatz, S. H. (1996). *Nonlinear dynamics and chaos*.
108. Szczepanik, M. (2011). Wpływ inflacji na wynagrodzenia w sektorze przedsiębiorstw w gospodarce polskiej. *Zeszyty Naukowe Wyższej Szkoły Bankowej we Wrocławiu*, (22), 7-22.
109. Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: the mother of all pandemics. *Revista Biomedica*, 17(1), 69-79.
110. Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, O., Lestari, H., Bahar, H., & Asfian, P. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *The Science of the total environment*, 725, 138436. <https://doi.org/10.1016/j.scitotenv.2020.138436>
111. Tsoy, D., Tirasawasdichai, T., & Kurpayanidi, K. I. (2021). Role of social media in shaping public risk perception during Covid-19 pandemic: a theoretical review. *International Journal of Management Science and Business Administration*, 7(2), 35-41.
112. Ülke, V., Sahin, A., & Subasi, A. (2018). A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. *Neural Computing and Applications*, 30, 1519-1527.
113. Varlık, N. (2021). Why Is Black Death Black? European Gothic Imaginaries of ‘Oriental’ Plague. In *Plague Image and Imagination from Medieval to Modern Times* (pp. 11-35). Palgrave Macmillan, Cham.
114. Van Solingen, R., Basili, V., Caldiera, G., & Rombach, H. D. (2002). Goal question metric (gqm) approach. *Encyclopedia of software engineering*.
115. Wagner, D. M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J. W., ... & Poinar, H. (2014). *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *The Lancet infectious diseases*, 14(4), 319-326.

116. Walkowiak, M. P., & Walkowiak, D. (2022). Underestimation in Reporting Excess COVID-19 Death Data in Poland during the First Three Pandemic Waves. *International journal of environmental research and public health*, 19(6), 3692.
117. Wang, Chen, Peter W. Horby, Frederick G. Hayden, and George F. Gao. "A novel coronavirus outbreak of global health concern." *The lancet* 395, no. 10223 (2020): 470-473.
118. Werner, P. A., Kęsik-Brodacka, M., Nowak, K., Olszewski, R., Kaleta, M., & Liebers, D. T. (2022). Modeling the Spatial and Temporal Spread of COVID-19 in Poland Based on a Spatial Interaction Model. *ISPRS International Journal of Geo-Information*, 11(3), 195.
119. Werner, P. A., Skrynyk, O., Porczek, M., Szczepankowska-Bednarek, U., Olszewski, R., & Kęsik-Brodacka, M. (2021). The Effects of Climate and Bioclimate on COVID-19 Cases in Poland. *Remote Sensing*, 13(23), 4946.
120. World Health Organization. (2020). Considerations for quarantine of individuals in the context of containment for coronavirus disease (COVID-19): interim guidance, 19 March 2020 (No. WHO/2019-nCoV/IHR\_Quarantine/2020.2). World Health Organization.
121. Wiliński, A., Kupracz, Ł., Senejko, A., & Chrzęstek, G. (2022). COVID-19: average time from infection to death in Poland, USA, India and Germany. *Quality & Quantity*, 1-18.
122. Wright, J. H. (2009). Forecasting US inflation by Bayesian model averaging. *Journal of Forecasting*, 28(2), 131-144.
123. Xie, J., & Zhu, Y. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. *Science of the Total Environment*, 724, 138201.
124. Zafar, R. F., Qayyum, A., & Ghouri, S. P. (2015). Forecasting inflation using functional time series analysis.
125. Zbarachewicz, B. (2020). Uczelnie w czasie koronawirusa: problemy transformacji cyfrowej i postulaty na przyszłość. *Wiedza Obronna*.
126. Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. *Chaos, Solitons & Fractals*, 140, 110121.

127. Zhao, Z., Nehil-Puleo, K., & Zhao, Y. (2020). How well can we forecast the COVID-19 pandemic with curve fitting and recurrent neural networks?. medRxiv.
128. Zietz, B. P., & Dunkelberg, H. (2004). The history of the plague and the research on the causative agent *Yersinia pestis*. *International journal of hygiene and environmental health*, 207(2), 165-178.
129. „About pandemic phases”, <https://www.euro.who.int/en/health-topics/communicable-diseases/influenza/data-and-statistics/pandemic-influenza/about-pandemic-phases> (2022). Stan witryny na dzień 02.06.2022 r.
130. Google LLC "Google COVID-19 Community Mobility Reports". <https://www.google.com/covid19/mobility/> (2022). Stan witryny na dzień 31.05.2022 r.
131. „Global HIV & AIDS statistics — Fact sheet” <https://www.unaids.org/en/resources/fact-sheet>. Stan witryny na dzień 18.05.2022 r.
132. „Commercial flights down 42% in 2020”, <https://www.flightradar24.com/blog/commercial-flights-down-42-in-2020/> (2021). Stan witryny na dzień 31.05.2022 r.
133. „Computing Classification System”, <https://dl.acm.org/ccs> 2023. Stan witryny na dzień 23.02.2023 r.
134. „COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)”, <https://gisanddata.maps.arcgis.com/apps/dashboards/bda7a594740fd40299423467b48e9ecf6> (2022). Stan witryny na dzień 31.05.2022 r.
135. „COVID-19: Google Mobility Trends”, <https://ourworldindata.org/covid-google-mobility-trends> (2022). Stan witryny na dzień 31.05.2022 r.
136. „Covid 19 - ICM UW” <https://covid-19.icm.edu.pl/>. Stan witryny na dzień 19.05.2022 r.
137. "COVID-19 Individual Risk Assessment” <https://crs19.pl/> (2022). Stan witryny na dzień 19.05.2022 r.
138. „covsirphy · PyPI”, <https://pypi.org/project/covsirphy/> (2022). Stan witryny na dzień 07.06.2022 r.

139. „Domo Resource - Data Never Sleeps 7.0”,  
<https://www.domo.com/learn/infographic/data-never-sleeps-7> (2020). Stan witryny na dzień 01.06.2022 r.
140. „Domo Resource - Data Never Sleeps 8.0”,  
<https://www.domo.com/learn/infographic/data-never-sleeps-8> (2021). Stan witryny na dzień 01.06.2022 r.
141. „Domo Resource - Data Never Sleeps 9.0”,  
<https://www.domo.com/learn/infographic/data-never-sleeps-9> (2022). Stan witryny na dzień 01.06.2022 r.
142. „Drop in aircraft observations could have impact on weather forecasts”,  
<https://www.ecmwf.int/en/about/media-centre/news/2020/drop-aircraft-observations-could-have-impact-weather-forecasts> (2020). Stan witryny na dzień 31.05.2022 r.
143. „Economic slowdown as a result of COVID is no substitute for Climate Action”,  
<https://public.wmo.int/en/media/news/economic-slowdown-result-of-covid-no-substitute-climate-action> (2020). Stan witryny na dzień 31.05.2022 r.
144. „Fit nonlinear regression model - MATLAB fitnlm”  
[https://www.mathworks.com/help/stats/fitnlm.html?searchHighlight=fitnlm%20&s\\_tid=srchtitle\\_support\\_results\\_1\\_fitnlm%20](https://www.mathworks.com/help/stats/fitnlm.html?searchHighlight=fitnlm%20&s_tid=srchtitle_support_results_1_fitnlm%20) 2023. Stan witryny na dzień 05.12.2023 r.
145. „Impacts on the Global Observing System”,  
<https://public.wmo.int/en/resources/coronavirus-covid-19/impacts-global-observing-system> (2020). Stan witryny na dzień 31.05.2022 r.
146. „MOCOS : Strona Główna” <https://mocos.pl/pl/> (2022). Stan witryny na dzień 19.05.2022
147. „numpy.linalg.pinv NumPy v1.23 Manual”,  
<https://numpy.org/doc/stable/reference/generated/numpy.linalg.pinv.html> (2022). Stan witryny na dzień 31.05.2022 r.
148. „Polynomial curve fitting - MATLAB polyfit”,  
[https://www.mathworks.com/help/matlab/ref/polyfit.html?searchHighlight=polyfit&s\\_tid=srchtitle\\_support\\_results\\_1\\_polyfit](https://www.mathworks.com/help/matlab/ref/polyfit.html?searchHighlight=polyfit&s_tid=srchtitle_support_results_1_polyfit) 2023. Stan witryny na dzień 05.12.2023 r.

149. "Polynomial evaluation - MATLAB polyval"  
[https://www.mathworks.com/help/matlab/ref/polyval.html?s\\_tid=doc\\_ta](https://www.mathworks.com/help/matlab/ref/polyval.html?s_tid=doc_ta) 2023. Stan witryny na dzień 05.12.2023 r.
150. "Prices - Inflation (CPI) - OECD Data" <https://data.oecd.org/price/inflation-cpi.htm> (2023). Stan witryny na dzień 1.10.2023 r.
151. „Smoothing Splines - MATLAB & Simulink”,  
<https://www.mathworks.com/help/curvefit/smoothing-splines.html> 2023. Stan witryny na dzień 05.12.2023 r.
152. „The recovery from the COVID-19 crisis must lead to a different economy”,  
<https://www.un.org/en/un-coronavirus-communications-team/launch-report-socio-economic-impacts-covid-19> (2020). Stan witryny na dzień 31.05.2022 r.
153. Tidy, J. (2021). Coronavirus: Israel enables emergency spy powers. 2020. URL <https://www.bbc.com/news/technology-51930681>. Stan witryny na dzień 14.06.2022 r.
154. „Zespół Ministra Zdrowia ds. monitorowania i prognoz - Koronawirus: informacje i zalecenia - Portal Gov.pl" <https://www.gov.pl/web/koronawirus/zespol-ministra-zdrowia-do-spraw-monitorowania-i-prognoz> (2022). Stan witryny na dzień 19.05.2022 r.

## Spis ilustracji

- Rys. 1. Miejsce niniejszej rozprawy doktorskiej w klasyfikacji ACM Computing Classification System. Źródło: „Computing Classification System”, <https://dl.acm.org/ccs> 2023. 10
- Rys. 2. Zmiana odwiedzin poszczególnych kategorii miejsc w Wielkiej Brytanii od początku pandemii. Graficzna interpretacja danych z Google COVID-19 Community Mobility Reports („COVID-19: Google Mobility Trends”, <https://ourworldindata.org/covid-google-mobility-trends> (2022)) 31
- Rys. 3. Zmiana czasu spędzanego (świat z podziałem na kraje) w domu od wybuchu pandemii do 15 kwietnia 2020 roku w stosunku do czasu sprzed pandemii. Graficzna interpretacja danych z Google COVID-19 Community Mobility Reports („COVID-19: Google Mobility Trends”, <https://ourworldindata.org/covid-google-mobility-trends> (2022)). .....32
- Rys. 4. Zobrazowanie zróżnicowania światowego podejścia do kontroli sanitarnej dla osób wjeżdżających do danego kraju. Graficzna interpretacja danych z Google COVID-19 Community Mobility Reports („COVID-19: Google Mobility Trends”, <https://ourworldindata.org/covid-google-mobility-trends> (2022)). .....33
- Rys. 5. Porównanie wyników badań predykcji zakażeń za pomocą modeli SIR (czarny słupek) i SEIR (szary słupek) na podstawie RMSE. Źródło: Alenezi et al. (2021). .....42
- Rys. 6. Porównanie wyników badań predykcji ozdowień za pomocą modeli SIR (czarny słupek) i SEIR (szary słupek) na podstawie RMSE. Źródło: Alenezi et al. (2021). .....43
- Rys. 7. Model epidemiologiczny SEIR (zaznaczona część) z modyfikacją zaproponowaną przez Peng et al. (2020). Źródło: Peng et al. (2020). .....44
- Rys. 8. Podział na typ agentów oraz zależności ekonomiczne pomiędzy agentami w modelu proponowanym przez Silva et al. (2020). Źródło: Silva et al. (2020). .....46
- Rys. 9. Analiza zakażeń dla poszczególnych scenariuszy modelu Silva et al. (2020). Źródło: Silva et al. (2020). .....47
- Rys. 10. Struktura drzewiasta obrazująca scenariusz rozprzestrzeniania się COVID-19 w zaproponowanym przez Bhatnagar (2020) modelu predykcji. Źródło: Bhatnagar (2020) .....48
- Rys. 11. Rozprzestrzenianie się COVID-19 we Włoszech – sumaryczna dzienna liczba zakażeń. Kolorem niebieskim oznaczono prognozę w modelu bez ograniczeń, kolorem różowym

prognozę w modelu z ograniczeniami, natomiast kolorem brunatnym dane aktualne na podstawie danych z portalu worldometer.info. Źródło: Bhatnagar (2020) .....	50
Rys. 12. Rozprzestrzenianie się COVID-19 w USA – sumaryczna dzienna liczba zakażeń. Kolorem niebieskim oznaczono prognozę w modelu bez ograniczeń, kolorem różowym prognozę w modelu z ograniczeniami, natomiast kolorem brunatnym dane aktualne na podstawie danych z portalu worldometer.info. Źródło: Bhatnagar (2020) .....	51
Rys. 13. Prognoza szeregu czasowego zakażeń COVID-19 w Indiach za pomocą metody opartej na ARIMA zaproponowanej przez Sharma et al.(2020) Linia ciągła koloru niebieskiego przedstawia dane wykorzystane do trenowania modelu, linia przerywana koloru czarnego oznacza dane prognozowane, linia niebieska przerywana oznacza dane aktualne, natomiast czerwone linie oznaczają RMSE z 95% przedziałem pewności. Źródło: (Sharma et al.(2020)). .....	54
Rys. 14. Zaobserwowana i przewidywana liczba nowych przypadków COVID-19 w Iranie z podziałem szeregu czasowego na dane do trenowania modelu (do 24 marca 2020 r.) i do ewaluacji modeli ANN i ARIMA (od 25 marca 2020 r. do 31 marca 2020 r.). Niebieska ciągła linia oznacza dane rzeczywiste, zielona przerywana linia to predykcja za pomocą ANN, natomiast granatowa przerywana linia to predykcja za pomocą ARIMA. Źródło: Moftakhar et al. (2020).....	56
Rys. 15. Wykres pudełkowy wartości inflacji ze 100 ostatnich miesięcy w poszczególnych krajach. Stan na październik 2023. (Źródło: <a href="https://data.oecd.org/price/inflation-cpi.htm">https://data.oecd.org/price/inflation-cpi.htm</a> ).....	65
Rys. 16. Wykres pudełkowy wartości inflacji (z pominięciem outlierów) ze 100 ostatnich miesięcy w poszczególnych krajach. Stan na październik 2023. (Źródło: <a href="https://data.oecd.org/price/inflation-cpi.htm">https://data.oecd.org/price/inflation-cpi.htm</a> ).....	67
Rys. 17. Poziom inflacji w Polsce w okresie 1.2006 r. – 9.2023 r.(Źródło: opracowanie własne na podstawie <a href="https://data.oecd.org/price/inflation-cpi.htm">https://data.oecd.org/price/inflation-cpi.htm</a> ).....	68
Rys. 18. Podział wybranego zakresu dat na 7 dniowe okna czasowe. Źródło: opracowania własne. 88	
Rys. 19. Podział okna czasowego na 5 binów. Źródło: opracowania własne. ....	89
Rys. 20. Zobrazowanie przesunięcia kolejnego okna czasowego ( $tw=2$ ) w celu uzyskania informacji o przejściu pomiędzy binami (tutaj informacja z dnia 03.09.2021 r.). Źródło: opracowania własne.....	90
Rys. 21. Zobrazowany przykład określania przejścia dla 2 kolejnych stanów (początkowego i końcowego). Źródło: opracowanie własne.....	91

Rys. 22. Zobrazowana klasyfikacja stanów i określenie przejść w oknie czasowym. Źródło: opracowanie własne. ....	92
Rys. 23. Przykładowa macierz przejść dla danego szeregu czasowego wraz z objaśnieniem stosowanej notacji. Źródło: opracowania własne. ....	92
Rys. 24. Zobrazowana macierz prawdopodobieństw P w formie mapy ciepła (zielone odcienie przyjmują wartości wyższe, żółte wartości bliskie medianie oraz czerwone minimalne). Pogrubioną czcionką oznaczono wartości maksymalne w danych wierszach. Źródło: opracowania własne. ....	93
Rys. 25. Zobrazowanie przejść pomiędzy binami na podstawie wartości maksymalnych prawdopodobieństwa w macierzy prawdopodobieństw Pogrubioną czcionką oznaczono wartości maksymalne w danych wierszach. Źródło: opracowania własne. ....	93
Rys. 26. Otrzymane wartości kolumn macierzy <i>Confnorm</i> . Przebiegi dla poszczególnych krajów. Źródło: opracowania własne. ....	100
Rys. 27. Wybrane kNN-najbliższe przebiegi względem kraju bazowego (oznaczone kolorem czerwonym) oraz przebieg kraju bazowego (oznaczony kolorem zielonym). Źródło: opracowanie własne. ....	104
Rys. 28. Wykres pudełkowy przedstawiający w sposób graficzny reprezentację podstawowej analizy statystycznej kraju bazowego (tutaj Polska) oraz n-najbliższych sąsiadów. Pierwszy kraj to kraj bazowy, kolejne to jego najbliżsi sąsiedzi. Na przykładzie kraju bazowego widać 2 outliery reprezentowane w postaci czerwonych punktów. ....	106
Rys. 29. Wykres rozkładów dwuwymiarowych par w zbiorze danych reprezentujących kraj bazowy i n-najbliższych sąsiadów. ....	107
Rys. 30. Mapa ciepła obrazująca korelację wzajemną między krajem bazowym oraz n-najbliższymi sąsiadami. Źródło: opracowanie własne. ....	108
Rys. 31. Wykres przebiegu średniej kroczącej MA <sub>7</sub> dla każdego z n-najbliższych sąsiadów dla kraju bazowego. Źródło: opracowanie własne. ....	111
Rys. 32. Krzywa referencyjna (oznaczona kolorem zielonym) zestawiona z średnią krocząca kraju bazowego (oznaczona kolorem czerwonym) w przedziale uczenia. Źródło: opracowanie własne. ....	112
Rys. 33. Zobrazowane przebiegi krzywej referencyjnej i przebiegu przewidywanych wartości szeregu czasowego. Źródło: opracowanie własne. ....	113

Rys. 34. Przebieg wartości rzeczywistych (kolor zielony) oraz prognozowanych (kolor czerwony) dla kraju bazowego w horyzoncie predykcji  $h_p=14$  dni. Źródło: opracowanie własne.

114

Rys. 35. Wykres przebiegów APE (linia koloru zielonego) oraz MAPE (linia koloru niebieskiego) w horyzoncie predykcji  $h_p$ . Źródło: opracowanie własne..... 116

Rys. 36. Zobrazowanie ekstrapolacji regresji wielomianowej z uwzględnieniem parametru kwst – kroków wstecz. Niebieskie punkty połączone niebieską linią oznaczają wartości rzeczywiste. Niebieskie kółka oznaczają liczbę kroków wstecz (wartość parametru kwst). Czerwona linia to to wielomian n-tego stopnia i jego ekstrapolowana część oznaczona zieloną przerywaną linią do punkty oznaczonego X w kolorze magenta – prognozowana wartość w  $h_p=1$ . Źródło: opracowanie własne. 120

Rys. 37. Zobrazowanie graficzne korygowania wartości prognozowanych za pomocą korekty ECM. Niebieska linia oraz punkty oznaczają wartości rzeczywiste, czerwona linia oraz czerwone punkty to pierwotna predykcja – z wcześniejszych kroków, zielona linia i punkty to skorygowane wartości poprzedniej predykcji. Czerwony diament oznacza przewidywaną wartość w  $h_p=1$ , natomiast zielona gwiazda oznacza skorygowaną wartość przewidywaną na podstawie błędów z poprzednich kroków. Źródło: opracowanie własne..... 122

Rys. 38. Zobrazowanie interpolacji splajnu. Punkty oznaczone niebieskimi kółkami oznaczają wartości rzeczywiste szeregu czasowego, czerwona linia to przebieg splajnu. Źródło: opracowanie własne. 125

Rys. 39. Demonstracja predykcji regresją wygładzającą splajnem. Niebieskie punkty połączone linią to wartości rzeczywiste, czerwona linia to interpolowany splajn, zielona linia to ekstrapolacja splajnu, czerwona gwiazdka oznacza prognozowaną wartość w  $h_p=1$ . Źródło: opracowanie własne..... 126

Rys. 40. Wykres przebiegu czasowego skumulowanej dobowej liczby zakażeń w Islandii od 1 sierpnia do 15 września 2021 r. Źródło: opracowanie własne. .... 135

Rys. 41. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji  $h_p$ : rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Islandii. Oś rzędnych zawiera kolejne dni horyzontu predykcji. Źródło: opracowanie własne. .... 141

Rys. 42. Przebieg APE dla badanego przypadku - Islandii. Źródło: opracowanie własne. .... 142

Rys. 43. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji hp: rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Ugandy. Oś rzędnych zawiera kolejne dni horyzontu predykcji. Źródło: opracowanie własne. ....	143
Rys. 44. Przebieg APE dla badanego przypadku - Ugandy. Źródło: opracowanie własne. ...	143
Rys. 45. Przebieg APE dla badanego przypadku – Ugandy, dla zakresu dat: 22-28 luty 2022 r. Źródło: opracowanie własne.....	145
Rys. 46. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji hp: rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Ugandy dla zakresu dat: 22-28 luty 2022 r. Oś rzędnych zawiera kolejne dni horyzontu predykcji. Źródło: opracowanie własne. ....	145
Rys. 47. Przebieg APE dla badanego przypadku – Islandii, dla zakresu dat: 22-28 luty 2022 r. Źródło: opracowanie własne.....	146
Rys. 48. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji hp: rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Islandii dla zakresu dat: 22-28 luty 2022 r. Oś rzędnych zawiera kolejne dni horyzontu predykcji. Źródło: opracowanie własne. ....	147
Rys. 49. Skumulowana liczba zakażeń w Niemczech obejmująca badany okres. Źródło: opracowanie własne. ....	148
Rys. 50. Przebiegi znormalizowanych szeregów czasowych dobowych skumulowanych zakażeń dla wszystkich krajów z dostępnymi danymi. Źródło: opracowanie własne.....	149
Rys. 51. Mapa ciepła obrazująca korelację wzajemną między krajem bazowym oraz n-najbliższymi sąsiadami. Źródło: opracowanie własne.....	150
Rys. 52. Wykres rozkładów dwuwymiarowych par w zbiorze danych reprezentujących kraj bazowy i n-najbliższych sąsiadów. Źródło: opracowanie własne. ....	151
Rys. 53. Wykres pudełkowy przedstawiający w sposób graficzny reprezentację podstawowej analizy statystycznej kraju bazowego (tutaj Niemcy) oraz n-najbliższych sąsiadów. Pierwszy kraj to kraj bazowy, kolejne to jego najbliżsi sąsiedzi. Źródło: opracowanie własne. ....	152
Rys. 54. Rzeczywiste znormalizowane przebiegi czasowe skumulowanej dobowej liczby zakażeń dla n-najbliższych sąsiadów (oznaczone kolorem czerwonym) oraz kraju referencyjnego (oznaczonego kolorem zielonym). Źródło: opracowanie własne. ....	153

Rys. 55. Średnia krocząca MA znormalizowanych przebiegów szeregów czasowych skumulowanej dobowej liczby zakażeń dla n-najbliższych sąsiadów kraju referencyjnego. Źródło: opracowanie własne. ....	154
Rys. 56. Wykres przebiegów uśrednionej średniej kroczącej dla n-sąsiadów dobowej skumulowanej liczby zakażeń (oznaczona kolorem czerwonym) oraz średniej kroczącej skumulowanej dobowej liczby zakażeń dla kraju referencyjnego (oznaczona kolorem zielonym). Źródło: opracowanie własne. ....	155
Rys. 57. Przebieg rzeczywisty (kolor zielony) oraz prognoza w horyzoncie predykcji (kolor niebieski) w symulacji o parametrach $n=5$ i $hp=3$ w badanym okresie czasu dla Niemiec. Źródło: opracowanie własne. ....	156
Rys. 58. Przebieg szeregów czasowych skumulowanych dobowych zakażeń w horyzoncie predykcji $hp$ : rzeczywisty (kolor zielony) oraz prognozowany (kolor czerwony) dla Niemiec i parametrów $n=5$ i $hp=3$ . Źródło: opracowanie własne. ....	157
Rys. 59. Przebieg APE (kolor zielony) oraz MAPE (kolor niebieski) dla badanego przypadku – Niemiec, parametry $n=5$ i $hp=3$ . Źródło: opracowanie własne. ....	157
Rys. 60. Przebieg rzeczywisty (kolor zielony) oraz prognoza w horyzoncie predykcji (kolor niebieski) w symulacji o parametrach $n=5$ i $hp=21$ w badanym okresie czasu dla Niemiec. Źródło: opracowanie własne. ....	158
Rys. 61. Przebieg rzeczywisty (kolor zielony) oraz prognoza w horyzoncie predykcji (kolor niebieski) w symulacji o parametrach $n=5$ i $hp=21$ w badanym okresie czasu dla Niemiec. Źródło: opracowanie własne. ....	159
Rys. 62. Przebieg APE (kolor zielony) oraz MAPE (kolor niebieski) dla badanego przypadku – Niemiec, parametry $n=5$ i $hp=21$ . Źródło: opracowanie własne. ....	160
Rys. 63. Wykres szeregów czasowych dziennych przyrostów zakażeń w wybranych krajach. Czerwona ramka oznacza badany okres dla wybranego kraju do badań. Źródło: opracowanie własne. ....	163
Rys. 64. Przebieg APE dla predykcji szeregu czasowego metodą multiregresji z pseudo-inwersją Moore'a-Penrose'a dla $hp=7$ . Wartość APE(0) odpowiada wartości MAPE dla $hp=1$ . Źródło: opracowanie własne. ....	164

Rys. 65. Wartości rzeczywiste (kolor zielony) i prognozowane (kolor czerwony) szeregów czasowych - predykcja metodą multiregresji z pseudo-inwersją Moore'a-Penrose'a dla $hp=7$ . Źródło: opracowanie własne.....	165
Rys. 66. Przebieg APE dla predykcji szeregu czasowego metodą multiregresji z pseudo-inwersją Moore'a-Penrose'a dla $hp=7$ . Źródło: opracowanie własne. ....	167
Rys. 67. Wartości rzeczywiste (kolor zielony) i prognozowane (kolor czerwony) szeregów czasowych - predykcja metodą multiregresji z pseudo-inwersją Moore'a-Penrose'a dla $hp=21$ . Źródło: opracowanie własne.....	167
Rys. 68. Przebieg APE (kolor zielony) i MAPE (kolor czerwony) dla predykcji szeregu czasowego metodą n-najbliższych sąsiadów dla $hp=21$ . Źródło: opracowanie własne.....	171
Rys. 69. Przebieg rzeczywisty (kolor zielony) oraz prognoza w horyzoncie predykcji (kolor niebieski)szeregów czasowych skumulowanych dobowych zakażeń - predykcja metodą n-najbliższych sąsiadów dla $hp=21$ . Źródło: opracowanie własne. ....	172
Rys. 70. Wartości rzeczywiste (kolor zielony) i prognozowane (kolor czerwony) szeregów czasowych skumulowanych dobowych zakażeń - predykcja metodą n-najbliższych sąsiadów dla $hp=21$ . Źródło: opracowanie własne.....	172
Rys. 71. Przebieg skumulowanej liczby zakażeń w Indiach w zakresie dat uczenia z podziałem na tygodniowe okna. Źródło: opracowanie własne. ....	173
Rys. 72. Podział pierwszego tygodniowego okna czasowego badanego szeregu czasowego skumulowanych dobowych zakażeń w Indiach na 5 binów. Źródło: opracowanie własne.....	174
Rys. 73. Przebieg APE (kolor zielony) i MAPE (kolor czerwony) w badanym okresie predykcji dla Indii. Źródło: opracowanie własne. ....	176
Rys. 74. Porównanie wartości rzeczywistych i predykcji dobowej liczby zakażeń dla Indii w badanym zakresie dat z uwzględnieniem ograniczeń metody. Kolorem zielonym oznaczono dane rzeczywiste, kolorem czerwonym wartości prognozowane. Źródło: opracowanie własne. ....	177
Rys. 75. Porównanie badanych metod pod względem dokładności predykcji szeregów czasowych w zależności od długości horyzontu predykcji $hp$ za pomocą kryterium MAPE. Źródło: opracowanie własne. ....	179
Rys. 76. Wykres szeregów czasowych skumulowanej liczby szczepień dla wybranych krajów. Źródło: opracowanie własne.....	181

Rys. 77.	Wykres szeregów czasowych skumulowanych dobowych zakażeń w wybranych krajach. Źródło: opracowanie własne. ....	183
Rys. 78.	Analiza graficzna danych i wyników badania metodą n-najbliższych sąsiadów dla kraju Zimbabwe. Źródło: opracowanie własne. ....	185
Rys. 79.	Przebiegi szeregów czasowych przyrostów dobowych zakażeń dla badanych krajów afrykańskich. Źródło: opracowanie własne. ....	186
Rys. 80.	Wykres szeregów czasowych skumulowanych dobowych szczepień dla badanych krajów afrykańskich. Źródło: opracowanie własne. ....	187
Rys. 81.	Wykres ilustrujący dokładność MAPE w predykcji badanymi metodami dla wybranych krajów. Źródło: opracowanie własne. ....	188
Rys. 82.	Wykres szeregu czasowego dobowych przyrostów zakażeń w Indiach z zaznaczonymi fazami pandemii uwzględnionymi w badaniach. Źródło: opracowanie własne. .	191
Rys. 83.	Porównanie przebiegów APE dla predykcji szeregu czasowego skumulowanej liczby zakażeń w Indiach z rozróżnieniem na poszczególne fazy – metoda multiregresji z pseudoinwersją Moore’a – Penrose’a. Źródło: opracowanie własne. ....	193
Rys. 84.	Liczna obecność outlierów w analizie danych okna uczenia dla 4 fazy fragmentu szeregu czasowego skumulowanych dobowych zakażeń w Indiach. Źródło: opracowanie własne.	194
Rys. 85.	Zestawienie przebiegów rzeczywistych (kolor zielony) i prognozowanych(kolor niebieski) dla predykcji szeregu czasowego skumulowanej liczby zakażeń w Indiach z rozróżnieniem na poszczególne fazy – metoda n-najbliższych sąsiadów. Źródło: opracowanie własne.	196
Rys. 86.	Porównanie przebiegów APE i MAPE dla predykcji szeregu czasowego skumulowanej liczby zakażeń w Indiach z rozróżnieniem na poszczególne fazy – metoda n-najbliższych sąsiadów. Źródło: opracowanie własne. ....	197
Rys. 87.	Zestawienie przebiegów rzeczywistych (kolor zielony) i prognozowanych(kolor niebieski) dla predykcji szeregu czasowego skumulowanej liczby zakażeń w Indiach z rozróżnieniem na poszczególne fazy – metoda n-najbliższych sąsiadów. Źródło: opracowanie własne.	198

Rys. 88. Porównanie 20 ostatnich wartości rzeczywistych i prognozowanych (metody regresji wielomianowej (kolor zielony) i wygładzania splajnem (kolor magenta) dla USA. Źródło: opracowanie własne. ....	208
Rys. 89. Porównanie 75 ostatnich wartości rzeczywistych i prognozowanych (metody regresji wygładzania splajnem bez optymalizacji (kolor czerwony) i wygładzania splajnem z korektą ECM (kolor magenta) dla Rosji. Źródło: opracowanie własne. ....	209
Rys. 90. Porównanie wpływu optymalizacji i korekty ECM na MSE dla zastosowanych metod predykcji na całym zbiorze danych wejściowych. Źródło: opracowanie własne. ....	210
Rys. 91. Porównanie 16 ostatnich wartości rzeczywistych i prognozowanych (metody regresji wygładzania splajnem z optymalizacją (kolor zielony) i optymalizacji połączonej z korektą ECM (kolor magenta) dla Turcji. Źródło: opracowanie własne. ....	211
Rys. 92. Porównanie 16 ostatnich wartości rzeczywistych i prognozowanych (metody regresji wielomianem bez korekty (kolor zielony) i z korektą ECM (kolor magenta) dla Turcji. Źródło: opracowanie własne. ....	212

## Spis tabel

Tab. 1.	Matryca przedstawiająca zaobserwowane oraz potencjalne wpływy COVID-19 na zmiany klimatu oraz środowiska. Czerwonym kolorem wskazano negatywne wpływy, zielonym pozytywne a szarym neutralne [Cheval et al. (2020)].	25
Tab. 2.	Porównanie dokładności predykcji ozdowień COVID-19 za pomocą wybranych modeli dla populacji Włoch. Źródło: Zeroual et al. (2020).	57
Tab. 3.	Porównanie MAPE dla wybranych modeli wariantów LSTM w predykcji zakażeń i zgonów w Indiach i USA. Źródło: Shastri et al. (2020).	58
Tab. 4.	Format danych pobranych z CSSE przed wstępnym przetworzeniem. Źródło: opracowanie własne.	76
Tab. 5.	Macierz $Conf^f$ przedstawiająca dane wybranych krajów pogrupowane w kolumnach. Źródło: opracowanie własne.	77
Tab. 6.	Macierz $C1$ przedstawiająca dane wybranych krajów pogrupowane w kolumnach. Źródło: opracowanie własne.	78
Tab. 7.	Tabela przedstawiająca wielkość populacji, liczbę wykonanych szczepień w dniu odpowiadającym indeksowi $beg$ macierzy $Vac$ (w tym przypadku 1 styczeń 2021 r.) oraz gęstość szczepień (kolumna $Vac\_Density$ ). Źródło: opracowanie własne.	81
Tab. 8.	Macierz $C$ ze zmienną objaśnianą jako wektor $Y$ w notacji Ivakhnenko (Ivakhnenko et al. 1976) ze wskazaniem horyzontu prognozy. Źródło: opracowanie własne.	82
Tab. 9.	Macierz $C$ zawierająca predyktory – wektory skumulowanej liczby zakażeń w krajach $C_k$ w oknie czasowym w pomiędzy datami $beg$ i $fin$ . Źródło: opracowanie własne.	82
Tab. 10.	Wektor $y$ reprezentujący państwo będące zmienną objaśnianą $Y$ , który należy przesunąć o horyzont prognozy $hp$ w celu zbudowania modelu i umożliwienia predykcji.	82
Tab. 11.	Macierz $C$ oraz konkatenowany wektor $1$ . Źródło: opracowanie własne.	83
Tab. 12.	Zobrazowanie szeregów czasowych w modelu. Macierz $C$ oznaczono kolorem pomarańczowym, natomiast wektor $y$ przesunięty o $hp=7$ kolorem żółtym. Źródło: opracowanie własne.	84
Tab. 13.	Zobrazowanie szeregów czasowych w modelu na etapie predykcji wektora $Y_m$ o horyzoncie predykcji $hp=7$ począwszy od daty 22.01.2021 (oznaczony kolorem zielonym). Macierz $C_y$ oznaczono kolorem pomarańczowym. Źródło: opracowanie własne.	85

Tab. 14. Przygotowanie danych wejściowych do algorytmu Macierz Conf zawierająca 24 dniowy segment danych (skumulowane dobowe przypadki zakażeń) poprzedzający horyzont predykcji. Źródło: opracowania własne.....	97
Tab. 15. Macierz <i>Confnorm</i> zawierająca znormalizowane w przedziale $\langle 0,1 \rangle$ skumulowane wartości dobowych przyrostów zakażeń. Dane z poszczególnych krajów zawarte są w osobnych kolumnach. Źródło: opracowania własne.....	99
Tab. 16. Macierz <i>Dist</i> zawierająca bezwzględne wartości odległości od dobowej wartości przebiegu kraju bazowego. Źródło: opracowanie własne. ....	102
Tab. 17. Tabela <i>Distsum</i> zawierająca odległości przebiegu szeregu czasowego danego kraju od przebiegu szeregu czasowego kraju bazowego. Źródło: opracowanie własne.....	103
Tab. 18. $N=5$ najbliższych krajów i ich odległości od kraju bazowego <i>Distsum(C)</i> . Źródło: opracowania własne. ....	104
Tab. 19. Analiza statystyczna segmentu danych dla $n$ -najbliższych krajów. Kolumna ‘std’ zawiera wartość średniego odchylenia elementów szeregu czasowego. Źródło: opracowania własne. ....	105
Tab. 20. Wartości średniej kroczącej o rozmiarze okna 7 dni dla $n$ -sąsiadujących krajów. Źródło: opracowania własne.....	110
Tab. 21. Przykład uśrednionych wartości dniowych umieszczonych w kolumnie ‘avg’. Źródło: opracowanie własne. ....	112
Tab. 22. Wynik symulacji zawierający prognozowane wartości (kolumna ‘avg’), przebieg rzeczywisty (kolumna ‘actual_cases_no’), rzeczywista liczba skumulowanych wartości przyrostów dobowych (kolumna ‘actual_cases’), APE (kolumna ‘ape’) oraz MAPE (kolumna ‘MAPE’). Źródło: opracowanie własne. ....	115
Tab. 23. Porównanie metod wykorzystanych w badaniu (+ oznacza spełnienie kryterium, - oznacza brak spełnienia kryterium) . Źródło: opracowanie własne.....	130
Tab. 24. Przypisanie metod badawczych do określonego (za pomocą podejścia GQM) pytania. ....	131
Tab. 25. Dane i parametry algorytmów dla wybranych badań. (Źródło: mpracowanie własne). ....	132
Tab. 26. Lista krajów o największej i najmniejszej gęstości szczepień na dzień 1 sierpnia 2021 r. Kolorem zielonym oznaczono kraj o wysokiej wartości gęstości szczepień -Islandię, natomiast	

kolorem czerwonym kraj o niskiej wartości gęstości szczepień – Ugandę. Źródło: opracowanie własne na podstawie CSSE JHU. ....	134
Tab. 27. Macierz $C$ z konkatenowanym wektorem $\mathbf{1}$ (zgodnie z opisem w Tab. 11. ) dla badanego przypadku. Źródło: opracowanie własne.....	136
Tab. 28. Wektor $y$ reprezentujący państwo będące zmienną objaśnianą $Y$ , który należy przesunąć o horyzont prognozy $hp = 7$ w celu zbudowania modelu i umożliwienia predykcji zgodnie z opisem w 0 Źródło: opracowanie własne.....	137
Tab. 29. Macierz $C +$ powstała poprzez pseudoinwersję Moore’a-Penrose’a macierzy $C$ zgodnie z (30). Źródło: opracowanie własne. ....	138
Tab. 30. Obliczony wektor $A$ współczynników równania regresyjnego pozwalający na wyznaczenie modelu $Y_m$ zgodnie z (30). Źródło: opracowanie własne.....	138
Tab. 31. Macierz $C_y$ powstała w analogiczny sposób jak macierz $C$ z tą różnicą, że indeks wierszy szeregu czasowego jest przesunięty w czasie w przód o liczbę wierszy odpowiadający wartości horyzontu predykcji $hp$ , w tym przypadku $hp = 7$ . Źródło: opracowanie własne. ....	139
Tab. 32. Wektor $Y_m$ (właściwie druga połowa wektora wynikowego zgodnie z Tab. 13. ) stanowiący prognozowane wartości szeregu czasowego w horyzoncie predykcji i będący wynikiem równania regresyjnego (31). Źródło: opracowanie własne.....	140
Tab. 33. Wartości APE i MAPE uzyskane w badaniu dla kraju - Islandia. Źródło: opracowanie własne. ....	141
Tab. 34. Wartości APE i MAPE uzyskane w badaniu dla kraju - Uganda. Źródło: opracowanie własne. ....	144
Tab. 35. Zestawienie wartości rzeczywistych i prognozowanych oraz APE i MAPE dla badania predykcji o parametrach $n=5$ i $hp=3$ dla Niemiec. Źródło: opracowanie własne.....	155
Tab. 36. Zbiorczy wynik MAPE przeprowadzonego badania metodą $n$ -najbliższych sąsiadów z uwzględnieniem zmiany liczby sąsiadów i horyzontu predykcji. Źródło: opracowanie własne.	160
Tab. 37. Procentowe porównanie wyników MAPE względem wyników dla optymalnej wartości $n=7$ . Źródło: opracowanie własne.....	161
Tab. 38. Wynik predykcji szeregów czasowych skumulowanych dobowych zakażeń metodą multiregresji z pseudo-inwersją Moore’a-Penrose’a dla $hp=21$ . Źródło: opracowanie własne..	166

Tab. 39.	Zbiorcze zestawienie wyników MAPE dla predykcji metodą multiregresji z pseudoinwersją Moore'a-Penrose'a. Źródło: opracowanie własne. ....	168
Tab. 40.	Zbiorcze zestawienie wyników MAPE dla predykcji metodą n-najbliższych sąsiadów. Źródło: opracowanie własne. ....	169
Tab. 41.	Wynik predykcji szeregów czasowych skumulowanych dobowych zakażeń metodą n-najbliższych sąsiadów dla $hp=21$ . Źródło: opracowanie własne. ....	170
Tab. 42.	Macierz przejść dla badanego szeregu czasowego. Źródło: opracowanie własne. ....	174
Tab. 43.	Zobrazowana macierz prawdopodobieństw P. Źródło: opracowanie własne. ....	175
Tab. 44.	Wyniki analizy badanego okresu metodą z wykorzystaniem łańcuchów Markowa. Kolumny oznaczają kolejno: data, wartości rzeczywiste, wartości prognozowane, wartość odpowiadająca medianie bina, numer okna czasowego, numer bina źródłowego, docelowy bin, prognozowany docelowy bin, APE, MAPE. Źródło: opracowanie własne. ....	175
Tab. 45.	Porównanie dokładności metod predykcji za pomocą MAPE w zależności od długości horyzontu predykcji. Źródło: opracowanie własne. ....	179
Tab. 46.	Zestawienie wyników badań na podstawie MAPE dla krajów położonych geograficznie blisko siebie. Źródło: opracowanie własne. ....	189
Tab. 47.	Zbiorcze zestawienie wyników MAPE z predykcji szeregów czasowych skumulowanych dobowych zakażeń w poszczególnych fazach pandemii w Indiach z rozróżnieniem na metodę predykcji. ....	199
Tab. 48.	Wybór krajów do badania na podstawie kryteriów oraz lista 4 krajów o najbardziej pożądanej wartości metryk. (Źródło: opracowanie własne.) ....	201
Tab. 49.	Zbiorcze zestawienie wartości metryki predykcji MSE dla wybranych krajów w postaci mapy ciepła. Źródło: opracowanie własne. ....	205
Tab. 50.	Zbiorcze zestawienie wartości metryki predykcji RMSE dla wybranych krajów w postaci mapy ciepła. Źródło: opracowanie własne. ....	206
Tab. 51.	Zbiorcze porównanie średnich wartości metryk prognozowania zastosowanymi metodami. (Źródło: opracowanie własne.) ....	207

# Zawartość załączonej płyty DVD

## Zawartość załączonej

/		
Dane		
COVID-19		
20231209-225540-confirmed.csv		Dane pozyskane z GitHub zapisane w plikach po wstępnym przetworzeniu w celu badań.
20231209-225540-confirmed.xlsx		
20231209-225540-deaths.csv		
20231209-225540-deaths.xlsx		
20231209-225540-matlab_analysis.csv		
20231209-225540-metrics_dashboard.xlsx		
20231209-225540-recovered.csv		
20231209-225540-recovered.xlsx		
20231209-225540-vac.csv		
20231209-225540-vac.xlsx		
Inflacja		
CPI_AUT.m		Dane pozyskane z OECD przetworzone na wejściowe pliki do analizy w MATLAB
CPI_BEL.m		
CPI_BRA.m		
CPI_CAN.m		
CPI_CHE.m		
CPI_CHL.m		
CPI_CHN.m		
CPI_COL.m		
CPI_CRI.m		
CPI_CZE.m		
CPI_DEU.m		
CPI_DNK.m		
CPI_ESP.m		
CPI_EST.m		
CPI_FIN.m		
CPI_FRA.m		
CPI_GBR.m		
CPI_GER.m		
CPI_GRC.m		
CPI_HUN.m		
CPI_IDN.m		
CPI_IND.m		
CPI_IRL.m		
CPI_ISR.m		
CPI_ITA.m		
CPI_JPN.m		
CPI_KOR.m		
CPI_LTU.m		
CPI_LUX.m		
CPI_LVA.m		
CPI_MEX.m		
CPI_NLD.m		
CPI_NOR.m		
CPI_POL.m		
CPI_PRT.m		
CPI_RUS.m		
CPI_SAU.m		
CPI_SVK.m		
CPI_SVN.m		
CPI_SWE.m		
CPI_TUR.m		
CPI_USA.m		
CPI_ZAF.m		
f558591f-f9b5-40d6-b624-f1501d131b76_Data.csv		Dane pozyskane bezpośrednio z witryny OECD
Kod źródłowy		
Predykcja CPI		
Analiza_plikow_wynikowych_inflacja.m		Skrypt do analizy wyników
Inflacja_v6_iteracja_badanie.m		Skrypt wykonujący badanie predykcji inflacji

<ul style="list-style-type: none"> <li>└─ Preprocess_data_shrink_dataset_collective.m</li> <li>└─ Wyniki_FINAL.zip</li> <li>└─ Predykcja SARS-CoV-2 <ul style="list-style-type: none"> <li>└─ covid-19-data.py</li> <li>└─ covid-19-knnMA.py</li> <li>└─ covid-19-markov.py</li> <li>└─ covid19-multiregresja NG-vaccinations.py</li> </ul> </li> <li>└─ Rozprawa doktorska <ul style="list-style-type: none"> <li>└─ pdf <ul style="list-style-type: none"> <li>└─ Rozprawa doktorska.pdf</li> </ul> </li> </ul> </li> <li>└─ Wyniki <ul style="list-style-type: none"> <li>└─ Inflacja <ul style="list-style-type: none"> <li>└─ prediction_metrics.csv</li> </ul> </li> <li>└─ Wyniki_FINAL.zip</li> </ul> </li> </ul>	<p>Skrypt wybierający kraje do analizy na podstawie dynamiki inflacji Skompresowane pliki wynikowe z badań</p> <p>Skrypt do eksploracji danych Skrypt do predykcji: n-najbliższych sąsiadów ze średnią kroczącą Skrypt do predykcji: łańcuchy Markova Skrypt do predykcji: multiregresji z pseudoinwersją Moore'a – Penrose'a</p> <p>Treść rozprawy doktorskiej w formacie PDF</p> <p>Zestawienie uzyskanych metryk dla predykcji inflacji zastosowanymi metodami Skompresowane pliki wynikowe z badań</p>
---	--